



FIELD OF INVENTION

5 The present invention relates to a mechanism for preventing or reducing the possibility of inducing thermal-based attacks on an integrated circuit with on-board clock generation circuitry and at least some form of security.

10 The invention has primarily been developed for use in a printer that uses a plurality of security chips to ensure that modifications to operating parameters can only be modified in an authorized manner, and will be described with reference to this application. However, it will be appreciated that the invention can be applied to other fields in which analogous problems are faced.

BACKGROUND OF INVENTION

15 Manufacturing a printhead that has relatively high resolution and print-speed raises a number of problems.

20 Difficulties in manufacturing pagewidth printheads of any substantial size arise due to the relatively small dimensions of standard silicon wafers that are used in printhead (or printhead module) manufacture. For example, if it is desired to make an 8 inch wide pagewidth printhead, only one such printhead can be laid out on a standard 8-inch wafer, since such wafers are circular in plan. Manufacturing a pagewidth printhead from two or more smaller modules can reduce this limitation to some extent, but raises other problems related to providing a joint between adjacent printhead modules that is precise enough to avoid 25 visible artifacts (which would typically take the form of noticeable lines) when the printhead is used. The problem is exacerbated in relatively high-resolution applications because of the tight tolerances dictated by the small spacing between nozzles.

30 The quality of a joint region between adjacent printhead modules relies on factors including a precision with which the abutting ends of each module can be manufactured, the accuracy with which they can be aligned when assembled into a single printhead, and other more practical factors such as management of ink channels behind the nozzles. It will be appreciated that the difficulties include relative vertical displacement of the printhead modules with respect to each other.

35 Whilst some of these issues may be dealt with by careful design and manufacture, the level of precision required renders it relatively expensive to manufacture printheads within the required tolerances. It would be desirable to provide a solution to one or more of the problems associated with precision manufacture and assembly of multiple printhead modules to form a printhead, and especially a pagewidth printhead.

40 In some cases, it is desirable to produce a number of different printhead module types or lengths on a substrate to maximise usage of the substrate's surface area. However, different sizes and types of modules will have different numbers and layouts of print nozzles, potentially including different horizontal

and vertical offsets. Where two or more modules are to be joined to form a single printhead, there is also the problem of dealing with different seam shapes between abutting ends of joined modules, which again may incorporate vertical or horizontal offsets between the modules. Printhead controllers are usually dedicated application specific integrated circuits (ASICs) designed for specific use with a single type of

5 printhead module, that is used by itself rather than with other modules. It would be desirable to provide a way in which different lengths and types of printhead modules could be accounted for using a single printer controller.

Printer controllers face other difficulties when two or more printhead modules are involved, especially if it

10 is desired to send dot data to each of the printheads directly (rather than via a single printhead connected to the controller). One concern is that data delivered to different length controllers at the same rate will cause the shorter of the modules to be ready for printing before any longer modules. Where there is little difference involved, the issue may not be of importance, but for large length differences, the result is that the bandwidth of a shared memory from which the dot data is supplied to the modules is effectively left
15 idle once one of the modules is full and the remaining module or modules is still being filled. It would be desirable to provide a way of improving memory bandwidth usage in a system comprising a plurality of printhead modules of uneven length.

In any printing system that includes multiple nozzles on a printhead or printhead module, there is the

20 possibility of one or more of the nozzles failing in the field, or being inoperative due to manufacturing defect. Given the relatively large size of a typical printhead module, it would be desirable to provide some form of compensation for one or more "dead" nozzles. Where the printhead also outputs fixative on a per-nozzle basis, it is also desirable that the fixative is provided in such a way that dead nozzles are compensated for.

25 A printer controller can take the form of an integrated circuit, comprising a processor and one or more peripheral hardware units for implementing specific data manipulation functions. A number of these units and the processor may need access to a common resource such as memory. One way of arbitrating between multiple access requests for a common resource is timeslot arbitration, in which access to the
30 resource is guaranteed to a particular requestor during a predetermined timeslot.

One difficulty with this arrangement lies in the fact that not all access requests make the same demands on the resource in terms of timing and latency. For example, a memory read requires that data be fetched from memory, which may take a number of cycles, whereas a memory write can commence

35 immediately. Timeslot arbitration does not take into account these differences, which may result in accesses being performed in a less efficient manner than might otherwise be the case. It would be desirable to provide a timeslot arbitration scheme that improved this efficiency as compared with prior art timeslot arbitration schemes.

40 Also of concern when allocating resources in a timeslot arbitration scheme is the fact that the priority of an access request may not be the same for all units. For example, it would be desirable to provide a timeslot arbitration scheme in which one requestor (typically the memory) is granted special priority such

that its requests are dealt with earlier than would be the case in the absence of such priority.

In systems that use a memory and cache, a cache miss (in which an attempt to load data or an instruction from a cache fails) results in a memory access followed by a cache update. It is often desirable when

5 updating the cache in this way to update data other than that which was actually missed. A typical example would be a cache miss for a byte resulting in an entire word or line of the cache associated with that byte being updated. However, this can have the effect of tying up bandwidth between the memory (or a memory manager) and the processor where the bandwidth is such that several cycles are required to transfer the entire word or line to the cache. It would be desirable to provide a mechanism for updating
10 a cache that improved cache update speed and/or efficiency.

Most integrated circuits an externally provided signal as (or to generate) a clock, often provided from a dedicated clock generation circuit. This is often due to the difficulties of providing an onboard clock that can operate at a speed that is predictable. Manufacturing tolerances of such on-board clock generation

15 circuitry can result in clock rates that vary by a factor of two, and operating temperatures can increase this margin by an additional factor of two. In some cases, the particular rate at which the clock operates is not of particular concern. However, where the integrated circuit will be writing to an internal circuit that is sensitive to the time over which a signal is provided, it may be undesirable to have the signal be applied for too long or short a time. For example, flash memory is sensitive to being written too for too
20 long a period. It would be desirable to provide a mechanism for adjusting a rate of an on-chip system clock to take into account the impact of manufacturing variations on clockspeed.

One form of attacking a secure chip is to induce (usually by increasing) a clock speed that takes the logic outside its rated operating frequency. One way of doing this is to reduce the temperature of the

25 integrated circuit, which can cause the clock to race. Above a certain frequency, some logic will start malfunctioning. In some cases, the malfunction can be such that information on the chip that would otherwise be secure may become available to an external connection. It would be desirable to protect an integrated circuit from such attacks.

30 In an integrated circuit comprising non-volatile memory, a power failure can result in unintentional behaviour. For example, if an address or data becomes unreliable due to falling voltage supplied to the circuit but there is still sufficient power to cause a write, incorrect data can be written. Even worse, the data (incorrect or not) could be written to the wrong memory. The problem is exacerbated with multi-word writes. It would be desirable to provide a mechanism for reducing or preventing spurious writes when
35 power to an integrated circuit is failing.

In an integrated circuit, it is often desirable to reduce unauthorised access to the contents of memory.

This is particularly the case where the memory includes a key or some other form of security information that allows the integrated circuit to communicate with another entity (such as another integrated circuit,
40 for example) in a secure manner. It would be particularly advantageous to prevent attacks involving direct probing of memory addresses by physically investigating the chip (as distinct from electronic or logical attacks via manipulation of signals and power supplied to the integrated circuit).

It is also desirable to provide an environment where the manufacturer of the integrated circuit (or some other authorised entity) can verify or authorize code to be run on an integrated circuit.

- 5 Another desideratum would be the ability of two or more entities, such as integrated circuits, to communicate with each other in a secure manner. It would also be desirable to provide a mechanism for secure communication between a first entity and a second entity, where the two entities, whilst capable of some form of secure communication, are not able to establish such communication between themselves.
- 10 In a system that uses resources (such as a printer, which uses inks) it may be desirable to monitor and update a record related to resource usage. Authenticating ink quality can be a major issue, since the attributes of inks used by a given printhead can be quite specific. Use of incorrect ink can result in anything from misfiring or poor performance to damage or destruction of the printhead. It would therefore be desirable to provide a system that enables authentication of the correct ink being used, as well as
- 15 providing various support systems secure enabling refilling of ink cartridges.

In a system that prevents unauthorized programs from being loaded onto or run on an integrated circuit, it can be laborious to allow developers of software to access the circuits during software development.

- Enabling access to integrated circuits of a particular type requires authenticating software with a relatively high-level key. Distributing the key for use by developers is inherently unsafe, since a single leak of the key outside the organization could endanger security of all chips that use a related key to authorize programs. Having a small number of people with high-security clearance available to authenticate programs for testing can be inconvenient, particularly in the case where frequent incremental changes in programs during development require testing. It would be desirable to provide a mechanism for allowing access to one or more integrated circuits without risking the security of other integrated circuits in a series of such integrated circuits.
- 20
- 25

- In symmetric key security, a message, denoted by M , is *plaintext*. The process of transforming M into *ciphertext* C , where the substance of M is hidden, is called *encryption*. The process of transforming C back into M is called *decryption*. Referring to the encryption function as E , and the decryption function as D , we have the following identities:
- 30

$$\begin{aligned} E[M] &= C \\ D[C] &= M \end{aligned}$$

Therefore the following identity is true:

35

$$D[E[M]] = M$$

A symmetric encryption algorithm is one where:

- 40 • the encryption function E relies on key K_1 ,
- the decryption function D relies on key K_2 ,
- K_2 can be derived from K_1 , and

- K_1 can be derived from K_2 .

In most symmetric algorithms, K_1 equals K_2 . However, even if K_1 does not equal K_2 , given that one key can be derived from the other, a single key K can suffice for the mathematical definition. Thus:

5

$$E_K[M] = C$$

$$D_K[C] = M$$

The security of these algorithms rests very much in the key K . Knowledge of K allows anyone to encrypt or decrypt. Consequently K must remain a secret for the duration of the value of M . For example, M may

10 be a wartime message "My current position is grid position 123-456". Once the war is over the value of M is greatly reduced, and if K is made public, the knowledge of the combat unit's position may be of no relevance whatsoever. The security of the particular symmetric algorithm is a function of two things: the strength of the algorithm and the length of the key.

15 An asymmetric encryption algorithm is one where:

- the encryption function E relies on key K_1 ,
- the decryption function D relies on key K_2 ,
- K_2 cannot be derived from K_1 in a reasonable amount of time, and
- K_1 cannot be derived from K_2 in a reasonable amount of time.

20

Thus:

$$E_{K_1}[M] = C$$

$$D_{K_2}[C] = M$$

These algorithms are also called *public-key* because one key K_1 can be made public. Thus anyone can

25 encrypt a message (using K_1) but only the person with the corresponding decryption key (K_2) can decrypt and thus read the message.

In most cases, the following identity also holds:

$$E_{K_2}[M] = C$$

$$D_{K_1}[C] = M$$

30

This identity is very important because it implies that anyone with the public key K_1 can see M and know that it came from the owner of K_2 . No-one else could have generated C because to do so would imply knowledge of K_2 . This gives rise to a different application, unrelated to encryption - digital signatures.

35 A number of public key cryptographic algorithms exist. Most are impractical to implement, and many generate a very large C for a given M or require enormous keys. Still others, while secure, are far too slow to be practical for several years. Because of this, many public key systems are hybrid - a public key mechanism is used to transmit a symmetric session key, and then the session key is used for the actual

messages.

All of the algorithms have a problem in terms of key selection. A random number is simply not secure enough. The two large primes p and q must be chosen carefully - there are certain weak combinations

5 that can be factored more easily (some of the weak keys can be tested for). But nonetheless, key selection is not a simple matter of randomly selecting 1024 bits for example. Consequently the key selection process must also be secure.

Symmetric and asymmetric schemes both suffer from a difficulty in allowing establishment of multiple

10 relationships between one entity and a two or more others, without the need to provide multiple sets of keys. For example, if a main entity wants to establish secure communications with two or more additional entities, it will need to maintain a different key for each of the additional entities. For practical reasons, it is desirable to avoid generating and storing large numbers of keys. To reduce key numbers, two or more of the entities may use the same key to communicate with the main entity. However, this means that the 15 main entity cannot be sure which of the entities it is communicating with. Similarly, messages from the main entity to one of the entities can be decrypted by any of the other entities with the same key. It would be desirable if a mechanism could be provided to allow secure communication between a main entity and one or more other entities that overcomes at least some of the shortcomings of prior art.

20 In a system where a first entity is capable of secure communication of some form, it may be desirable to establish a relationship with another entity without providing the other entity with any information related the first entity's security features. Typically, the security features might include a key or a cryptographic function. It would be desirable to provide a mechanism for enabling secure communications between a first and second entity when they do not share the requisite secret function, key or other relationship to 25 enable them to establish trust.

A number of other aspects, features, preferences and embodiments are disclosed in the Detailed Description of the Preferred Embodiment below.

30 SUMMARY OF INVENTION

In accordance with the invention, there is provided an integrated circuit including an on-board system clock, the integrated circuit including a clock filter configured to determine a temperature of the integrated circuit and to alter an output of the system clock based on the temperature.

35

Preferably, the clock filter is configured to alter the output of the system clock in the event the temperature is outside a predetermined temperature range.

40 More preferably, altering the output includes preventing the clock signal from reaching one or more logical circuits on the integrated circuit to which it would otherwise be applied.

It is particularly preferred that the predetermined temperature range is selected such that a temperature-

related speed of the system clock output that is not due to the clock filter is within a predetermined frequency range. It is desirable that the frequency range be within an operating frequency of some or all of the logic circuitry to which the system clock is supplied.

- 5 In the preferred form of the invention, the clock filter is configured to prevent the system clock from reaching some or all of the logic circuitry in the event the temperature falls below a predetermined level. This level is chosen to be high enough that race conditions, in which the clock speeds up to the point where logic circuitry behavior becomes unpredictable, are avoided.
- 10 BRIEF DESCRIPTION OF THE DRAWINGS
Preferred and other embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings, in which:
 - Figure 1 is an example of state machine notation
 - Figure 2 shows document data flow in a printer
- 15 Figure 3 is an example of a single printer controller (hereinafter "SoPEC") A4 simplex printer system
Figure 4 is an example of a dual SoPEC A4 duplex printer system
Figure 5 is an example of a dual SoPEC A3 simplex printer system
Figure 6 is an example of a quad SoPEC A3 duplex printer system
Figure 7 is an example of a SoPEC A4 simplex printing system with an extra SoPEC used as DRAM
- 20 storage
Figure 8 is an example of an A3 duplex printing system featuring four printing SoPECs
Figure 9 shows pages containing different numbers of bands
Figure 10 shows the contents of a page band
Figure 11 illustrates a page data path from host to SoPEC
- 25 Figure 12 shows a page structure
Figure 13 shows a SoPEC system top level partition
Figure 14 is a Pulse generator RTL diagram
Figure 15 shows a SoPEC clock relationship
Figure 16 shows a CPR block partition
- 30 Figure 17 shows reset deglitch logic
Figure 18 shows reset synchronizer logic
Figure 19 is a clock gate logic diagram
Figure 20 shows a PLL and Clock divider logic
Figure 21 shows clock filter
- 35 Figure 22 shows tamper detection line
Figure 23 shows an oversize nMOS transistor layout of Tamper Detection Line
Figure 24 shows a Tamper Detection Line
Figure 25 shows how Tamper Detection Lines cover the Noise Generator
Figure 26 shows a prior art FET Implementation of CMOS inverter
- 40 Figure 27 shows non-flashing CMOS
Figure 28 shows a high level block diagram of QA Chip
Figure 29 shows an analogue unit

- Figure 30 shows a serial bus protocol for trimming
- Figure 31 shows a block diagram of a trim unit
- Figure 32 shows a block diagram of a CPU of the QA chip
- Figure 33 shows block diagram of an MIU
- 5 Figure 34 shows a block diagram of memory components
- Figure 35 shows a first byte sent to an IOU
- Figure 36 shows a block diagram of the IOU
- Figure 37 shows a relationship between external SDa and SClk and generation of internal signals
- Figure 38 shows a ring oscillator block diagram
- 10 Figure 39 shows a system clock duty cycle

DETAILED DESCRIPTION OF PREFERRED AND OTHER EMBODIMENTS

It will be appreciated that the detailed description that follows takes the form of a highly detailed design of the invention, including supporting hardware and software. A high level of detailed disclosure is provided 15 to ensure that one skilled in the art will have ample guidance for implementing the invention.

20 Imperative phrases such as "must", "requires", "necessary" and "important" (and similar language) should be read as being indicative of being necessary only for the preferred embodiment actually being described. As such, unless the opposite is clear from the context, imperative wording should not be interpreted as such. Nothing in the detailed description is to be understood as limiting the scope of the invention, which is intended to be defined as widely as is defined in the accompanying claims.

25 Indications of expected rates, frequencies, costs, and other quantitative values are exemplary and estimated only, and are made in good faith. Nothing in this specification should be read as implying that a particular commercial embodiment is or will be capable of a particular performance level in any measurable area.

30 It will be appreciated that the principles, methods and hardware described throughout this document can be applied to other fields. Much of the security-related disclosure, for example, can be applied to many other fields that require secure communications between entities, and certainly has application far beyond the field of printers.

SYSTEM OVERVIEW

35 The preferred of the present invention is implemented in a printer using microelectromechanical systems (MEMS) printheads. The printer can receive data from, for example, a personal computer such as an IBM compatible PC or Apple computer. In other embodiments, the printer can receive data directly from, for example, a digital still or video camera. The particular choice of communication link is not important, and can be based, for example, on USB, Firewire, Bluetooth or any other wireless or hardwired communications protocol.

PRINT SYSTEM OVERVIEW

This document describes the SoPEC (Small office home office Print Engine Controller) ASIC (Application Specific Integrated Circuit) suitable for use in, for example, SoHo printer products. The SoPEC ASIC is

5 intended to be a low cost solution for bi-lithic printhead control, replacing the multichip solutions in larger more professional systems with a single chip. The increased cost competitiveness is achieved by integrating several systems such as a modified PEC1 printing pipeline, CPU control system, peripherals and memory sub-system onto one SoC ASIC, reducing component count and simplifying board design.

10 This section will give a general introduction to Memjet printing systems, introduce the components that make a bi-lithic printhead system, describe possible system architectures and show how several SoPECs can be used to achieve A3 and A4 duplex printing. The section “**SoPEC ASIC**” describes the SoC SoPEC ASIC, with subsections describing the CPU, DRAM and Print Engine Pipeline subsystems. Each section gives a detailed description of the blocks used and their

15 operation within the overall print system. The final section describes the bi-lithic printhead construction and associated implications to the system due to its makeup.

Bi-LITHIC PRINthead NOTATION

A bi-lithic based printhead is constructed from 2 printhead ICs of varying sizes. The notation M:N is used

20 to express the size relationship of each IC, where M specifies one printhead IC in inches and N specifies the remaining printhead IC in inches.

The ‘SoPEC/MoPEC Bilitic Printhead Reference’ document [10] contains a description of the bi-lithic printhead and related terminology.

25 **DEFINITIONS**

The following terms are used throughout this specification:

Bi-lithic printhead	Refers to printhead constructed from 2 printhead ICs
CPU	Refers to CPU core, caching system and MMU.
30 ISI-Bridge chip	A device with a high speed interface (such as USB2.0, Ethernet or IEEE1394) and one or more ISI interfaces. The ISI-Bridge would be the ISIMaster for each of the ISI buses it interfaces to.
ISIMaster	The ISIMaster is the only device allowed to initiate communication on the Inter Sopec Interface (ISI) bus. The ISIMaster interfaces with the host.
35 ISISlave	Multi-SoPEC systems will contain one or more ISISlave SoPECs connected to the ISI bus. ISISlaves can only respond to communication initiated by the ISIMaster.
LEON	Refers to the LEON CPU core.
LineSyncMaster	The LineSyncMaster device generates the line synchronisation pulse that all
40	SoPECs in the system must synchronise their line outputs to.
Multi-SoPEC	Refers to SoPEC based print system with multiple SoPEC devices
Netpage	Refers to page printed with tags (normally in infrared ink).
PEC1	Refers to Print Engine Controller version 1, precursor to SoPEC used to

		control printheads constructed from multiple angled printhead segments.
Printhead IC		Single MEMS IC used to construct bi-lithic printhead
PrintMaster		The PrintMaster device is responsible for coordinating all aspects of the print operation. There may only be one PrintMaster in a system.
5	QA Chip	Quality Assurance Chip
	Storage SoPEC	An ISISlave SoPEC used as a DRAM store and which does not print.
	Tag	Refers to pattern which encodes information about its position and orientation which allow it to be optically located and its data contents read.

ACRONYM AND ABBREVIATIONS

10 The following acronyms and abbreviations are used in this specification

CFU	Contone FIFO Unit	
CPU	Central Processing Unit	
DIU	DRAM Interface Unit	
DNC	Dead Nozzle Compensator	
15	DRAM	Dynamic Random Access Memory
	DWU	DotLine Writer Unit
	GPIO	General Purpose Input Output
	HCU	Halftoner Compositor Unit
	ICU	Interrupt Controller Unit
20	ISI	Inter SoPEC Interface
	LDB	Lossless Bi-level Decoder
	LLU	Line Loader Unit
	LSS	Low Speed Serial interface
	MEMS	Micro Electro Mechanical System
25	MMU	Memory Management Unit
	PCU	SoPEC Controller Unit
	PHI	PrintHead Interface
	PSS	Power Save Storage Unit
	RDU	Real-time Debug Unit
30	ROM	Read Only Memory
	SCB	Serial Communication Block
	SFU	Spot FIFO Unit
	SMG4	Silverbrook Modified Group 4.
	SoPEC	Small office home office Print Engine Controller
35	SRAM	Static Random Access Memory
	TE	Tag Encoder
	TFU	Tag FIFO Unit
	TIM	Timers Unit
	USB	Universal Serial Bus

40 PSEUDOCODE NOTATION

In general the pseudocode examples use C like statements with some exceptions.

Symbol and naming conventions used for pseudocode.

//	Comment
=	Assignment
==,!=,<,>	Operator equal, not equal, less than, greater than
+,-,*,/,%	Operator addition, subtraction, multiply, divide, modulus
5 &,<,>,<<,>>,<~	Bitwise AND, bitwise OR, bitwise exclusive OR, left shift, right shift, complement
AND,OR,NOT	Logical AND, Logical OR, Logical inversion
[XX:YY]	Array/vector specifier
{a, b, c}	Concatenation operation
++, --	Increment and decrement
10 Register and signal naming conventions	In general register naming uses the C style conventions with capitalization to denote word delimiters. Signals use RTL style notation where underscore denote word delimiters. There is a direct translation between both convention. For example the <i>CmdSourceFifo</i> register is equivalent to <i>cmd_source_fifo</i> signal.
15 STATE MACHINE NOTATION	State machines should be described using the pseudocode notation outlined above. State machine descriptions use the convention of <u>underline</u> to indicate the cause of a transition from one state to another and plain text (no underline) to indicate the effect of the transition i.e. signal transitions which occur when the new state is entered.
20 A sample state machine is shown in Figure 1.	Printing Considerations
25	A bi-lithic printhead produces 1600 dpi bi-level dots. On low-diffusion paper, each ejected drop forms a 22.5 μ m diameter dot. Dots are easily produced in isolation, allowing dispersed-dot dithering to be exploited to its fullest. Since the bi-lithic printhead is the width of the page and operates with a constant paper velocity, color planes are printed in perfect registration, allowing ideal dot-on-dot printing. Dot-on-dot printing minimizes 'muddying' of midtones caused by inter-color bleed.
30	A page layout may contain a mixture of images, graphics and text. Continuous-tone (contone) images and graphics are reproduced using a stochastic dispersed-dot dither. Unlike a clustered-dot (or amplitude-modulated) dither, a <i>dispersed-dot</i> (or frequency-modulated) dither reproduces high spatial frequencies (i.e. image detail) almost to the limits of the dot resolution, while simultaneously reproducing lower spatial frequencies to their full color depth, when spatially integrated by the eye. A <i>stochastic</i> dither matrix is carefully designed to be free of objectionable low-frequency patterns when tiled across the image. As such its size typically exceeds the minimum size required to support a particular number of intensity levels (e.g. 16x16x 8 bits for 257 intensity levels).
35	Human contrast sensitivity peaks at a spatial frequency of about 3 cycles per degree of visual field and then falls off logarithmically, decreasing by a factor of 100 beyond about 40 cycles per degree and becoming immeasurable beyond 60 cycles per degree [25][25]. At a normal viewing distance of 12 inches (about 300mm), this translates roughly to 200-300 cycles per inch (cpi) on the printed page, or 400-600 samples per inch according to Nyquist's theorem.
40	In practice, contone resolution above about 300 ppi is of limited utility outside special applications such as medical imaging. Offset printing of magazines, for example, uses contone resolutions in the range 150 to 300 ppi. Higher resolutions contribute slightly to color error through the dither.

Black text and graphics are reproduced directly using bi-level black dots, and are therefore not anti-aliased (i.e. low-pass filtered) before being printed. Text should therefore be *supersampled* beyond the perceptual limits discussed above, to produce smoother edges when spatially integrated by the eye. Text resolution up to about 1200 dpi continues to contribute to perceived text sharpness (assuming low-diffusion paper, of course).

5 A Netpage printer, for example, may use a contone resolution of 267 ppi (i.e. 1600 dpi / 6), and a black text and graphics resolution of 800 dpi. A high end office or departmental printer may use a contone resolution of 320 ppi (1600 dpi / 5) and a black text and graphics resolution of 1600 dpi. Both formats are capable of exceeding the quality of commercial (offset) printing and photographic reproduction.

10

Document Data Flow

CONSIDERATIONS

Because of the page-width nature of the bi-lithic printhead, each page must be printed at a constant speed to avoid creating visible artifacts. This means that the printing speed can't be varied to match the 15 input data rate. Document rasterization and document printing are therefore decoupled to ensure the printhead has a constant supply of data. A page is never printed until it is fully rasterized. This can be achieved by storing a compressed version of each rasterized page image in memory.

This decoupling also allows the RIP(s) to run ahead of the printer when rasterizing simple pages, buying time to rasterize more complex pages.

20

Because contone color images are reproduced by stochastic dithering, but black text and line graphics are reproduced directly using dots, the compressed page image format contains a separate foreground bi-level black layer and background contone color layer. The black layer is composited over the contone layer after the contone layer is dithered (although the contone layer has an optional black component). A final layer of Netpage tags (in infrared or black ink) is optionally added to the page for printout.

25

Figure 2 shows the flow of a document from computer system to printed page.

At 267 ppi for example, a A4 page (8.26 inches × 11.7 inches) of contone CMYK data has a size of 26.3MB. At 320 ppi, an A4 page of contone data has a size of 37.8MB. Using lossy contone compression algorithms such as JPEG [27], contone images compress with a ratio up to 10:1 without noticeable loss of quality, giving compressed page sizes of 2.63MB at 267 ppi and 3.78 MB at 320 ppi.

30 At 800 dpi, a A4 page of bi-level data has a size of 7.4MB. At 1600 dpi, a Letter page of bi-level data has a size of 29.5 MB. Coherent data such as text compresses very well. Using lossless bi-level compression algorithms such as SMG4 fax, ten-point plain text compresses with a ratio of about 50:1. Lossless bi-level compression across an average page is about 20:1 with 10:1 possible for pages which compress poorly. The requirement for SoPEC is to be able to print text at 10:1 compression. Assuming 10:1 compression 35 gives compressed page sizes of 0.74 MB at 800 dpi, and 2.95 MB at 1600 dpi.

Once dithered, a page of CMYK contone image data consists of 116MB of bi-level data. Using lossless bi-level compression algorithms on this data is pointless precisely because the optimal dither is stochastic - i.e. since it introduces hard-to-compress disorder.

Netpage tag data is optionally supplied with the page image. Rather than storing a compressed bi-level 40 data layer for the Netpage tags, the tag data is stored in its raw form. Each tag is supplied up to 120 bits of raw variable data (combined with up to 56 bits of raw fixed data) and covers up to a 6mm × 6mm area (at 1600 dpi). The absolute maximum number of tags on a A4 page is 15,540 when the tag is only 2mm ×

2mm (each tag is 126 dots \times 126 dots, for a total coverage of 148 tags \times 105 tags). 15,540 tags of 128 bits per tag gives a compressed tag page size of 0.24 MB.

The multi-layer compressed page image format therefore exploits the relative strengths of lossy JPEG contone image compression, lossless bi-level text compression, and tag encoding. The format is compact

5 enough to be storage-efficient, and simple enough to allow straightforward real-time expansion during printing.

Since text and images normally don't overlap, the normal worst-case page image size is image only, while the normal best-case page image size is text only. The addition of worst case Netpage tags adds 0.24MB to the page image size. The worst-case page image size is text over image plus tags. The

10 average page size assumes a quarter of an average page contains images. Table 1 shows data sizes for compressed Letter page for these different options.

Table 1. Data sizes for A4 page (8.26 inches \times 11.7 inches)

	267 ppi contone 800 dpi bi-level	320 ppi contone 1600 dpi bi-level
Image only (contone), 10:1 compression	2.63 MB	3.78 MB
Text only (bi-level), 10:1 compression	0.74 MB	2.95 MB
Netpage tags, 1600 dpi	0.24 MB	0.24 MB
Worst case (text + image + tags)	3.61 MB	6.67 MB
Average (text + 25% image + tags)	1.64 MB	4.25 MB

15 DOCUMENT DATA FLOW

The Host PC rasterizes and compresses the incoming document on a page by page basis. The page is restructured into bands with one or more bands used to construct a page. The compressed data is then transferred to the SoPEC device via the USB link. A complete band is stored in SoPEC embedded memory. Once the band transfer is complete the SoPEC device reads the compressed data, expands the 20 band, normalizes contone, bi-level and tag data to 1600 dpi and transfers the resultant calculated dots to the bi-lithic printhead.

The document data flow is

- The RIP software rasterizes each page description and compress the rasterized page image.
- The infrared layer of the printed page optionally contains encoded Netpage [5] tags at a 25 programmable density.
- The compressed page image is transferred to the SoPEC device via the USB normally on a band by band basis.
- The print engine takes the compressed page image and starts the page expansion.
- The first stage page expansion consists of 3 operations performed in parallel
- 30 • expansion of the JPEG-compressed contone layer
- expansion of the SMG4 fax compressed bi-level layer
- encoding and rendering of the bi-level tag data.
- The second stage dithers the contone layer using a programmable dither matrix, producing up to four bi-level layers at full-resolution.

- The second stage then composites the bi-level tag data layer, the bi-level SMG4 fax de-compressed layer and up to four bi-level JPEG de-compressed layers into the full-resolution page image.
- A fixative layer is also generated as required.

5 • The last stage formats and prints the bi-level data through the bi-lithic printhead via the printhead interface.

The SoPEC device can print a full resolution page with 6 color planes. Each of the color planes can be generated from compressed data through any channel (either JPEG compressed, bi-level SMG4 fax compressed, tag data generated, or fixative channel created) with a maximum number of 6 data channels

10 from page RIP to bi-lithic printhead color planes.

The mapping of data channels to color planes is programmable, this allows for multiple color planes in the printhead to map to the same data channel to provide for redundancy in the printhead to assist dead nozzle compensation.

15 Also a data channel could be used to gate data from another data channel. For example in stencil mode, data from the bilevel data channel at 1600 dpi can be used to filter the contone data channel at 320 dpi, giving the effect of 1600 dpi contone image.

PAGE CONSIDERATIONS DUE TO SoPEC

The SoPEC device typically stores a complete page of document data on chip. The amount of storage available for compressed pages is limited to 2Mbytes, imposing a fixed maximum on compressed page

20 size. A comparison of the compressed image sizes in Table 2 indicates that SoPEC would not be capable of printing worst case pages unless they are split into bands and printing commences before all the bands for the page have been downloaded. The page sizes in the table are shown for comparison purposes and would be considered reasonable for a professional level printing system. The SoPEC device is aimed at the consumer level and would not be required to print pages of that complexity. Target

25 document types for the SoPEC device are shown Table 2.

Table 2. Page content targets for SoPEC

Page Content Description	Calculation	Size (MByte)
Best Case picture Image, 267ppi with 3 colors, A4 size	8.26x11.7x267x267x3 @10:1	1.97
Full page text, 800dpi A4 size	8.26x11.7x800x800 @10:1	0.74
Mixed Graphics and Text - Image of 6 inches x 4 inches @ 267 ppi and 3 colors - Remaining area text ~73 inches ² , 800 dpi	6x4x267x267x3 @ 5:1 800x800x73 @ 10:1	1.55
Best Case Photo, 3 Colors, 6.6 MegaPixel Image	6.6 Mpixel @ 10:1	2.00

30 If a document with more complex pages is required, the page RIP software in the host PC can determine that there is insufficient memory storage in the SoPEC for that document. In such cases the RIP software

can take two courses of action. It can increase the compression ratio until the compressed page size will fit in the SoPEC device, at the expense of document quality, or divide the page into bands and allow SoPEC to begin printing a page band before all bands for that page are downloaded. Once SoPEC starts printing a page it cannot stop, if SoPEC consumes compressed data faster than the bands can be

5 downloaded a buffer underrun error could occur causing the print to fail. A buffer underrun occurs if a line synchronisation pulse is received before a line of data has been transferred to the printhead. Other options which can be considered if the page does not fit completely into the compressed page store are to slow the printing or to use multiple SoPECs to print parts of the page. A Storage SoPEC could be added to the system to provide guaranteed bandwidth data delivery. The print system could also be

10 constructed using an ISI-Bridge chip to provide guaranteed data delivery.

Memjet Printer Architecture

The SoPEC device can be used in several printer configurations and architectures.

In the general sense every SoPEC based printer architecture will contain:

- One or more SoPEC devices.
- One or more bi-lithic printheads.
- Two or more LSS busses.
- Two or more QA chips.
- USB 1.1 connection to host or ISI connection to Bridge Chip.
- ISI bus connection between SoPECs (when multiple SoPECs are used).

20 SYSTEM COMPONENTS

SoPEC Print Engine Controller

The SoPEC device contains several system on a chip (SoC) components, as well as the print engine pipeline control application specific logic.

Print Engine Pipeline (PEP) Logic

25 The PEP reads compressed page store data from the embedded memory, optionally decompresses the data and formats it for sending to the printhead. The print engine pipeline functionality includes expanding the page image, dithering the contone layer, compositing the black layer over the contone layer, rendering of Netpage tags, compensation for dead nozzles in the printhead, and sending the resultant image to the bi-lithic printhead.

30 *Embedded CPU*

SoPEC contains an embedded CPU for general purpose system configuration and management. The CPU performs page and band header processing, motor control and sensor monitoring (via the GPIO) and other system control functions. The CPU can perform buffer management or report buffer status to the host. The CPU can optionally run vendor application specific code for general print control such as

35 paper ready monitoring and LED status update.

Embedded Memory Buffer

A 2.5Mbyte embedded memory buffer is integrated onto the SoPEC device, of which approximately 2Mbytes are available for compressed page store data. A compressed page is divided into one or more bands, with a number of bands stored in memory. As a band of the page is consumed by the PEP for printing a new band can be downloaded. The new band may be for the current page or the next page.

Using banding it is possible to begin printing a page before the complete compressed page is downloaded, but care must be taken to ensure that data is always available for printing or a buffer underrun may occur.

An Storage SoPEC acting as a memory buffer or an ISI-Bridge chip with attached DRAM could be used

5 to provide guaranteed data delivery.

Embedded USB 1.1 Device

The embedded USB 1.1 device accepts compressed page data and control commands from the host PC, and facilitates the data transfer to either embedded memory or to another SoPEC device in multi-SoPEC systems.

10 Bi-lithic Printhead

The printhead is constructed by abutting 2 printhead ICs together. The printhead ICs can vary in size from 2 inches to 8 inches, so to produce an A4 printhead several combinations are possible. For example two printhead ICs of 7 inches and 3 inches could be used to create a A4 printhead (the notation is 7:3). Similarly 6 and 4 combination (6:4), or 5:5 combination. For an A3 printhead it can be constructed from 8:6 or 15 an 7:7 printhead IC combination. For photographic printing smaller printheads can be constructed.

LSS interface bus

Each SoPEC device has 2 LSS system buses for communication with QA devices for system authentication and ink usage accounting. The number of QA devices per bus and their position in the system is unrestricted with the exception that *PRINTER_QA* and *INK_QA* devices should be on separate LSS

20 busses.

QA devices

Each SoPEC system can have several QA devices. Normally each printing SoPEC will have an associated *PRINTER_QA*. Ink cartridges will contain an *INK_QA* chip. *PRINTER_QA* and *INK_QA* devices should be on separate LSS busses. All QA chips in the system are physically identical with flash 25 memory contents defining *PRINTER_QA* from *INK_QA* chip.

ISI interface

The Inter-SoPEC Interface (ISI) provides a communication channel between SoPECs in a multi-SoPEC system. The ISIMaster can be SoPEC device or an ISI-Bridge chip depending on the printer configuration. Both compressed data and control commands are transferred via the interface.

30 ISI-Bridge Chip

A device, other than a SoPEC with a USB connection, which provides print data to a number of slave SoPECs. A bridge chip will typically have a high bandwidth connection, such as USB2.0, Ethernet or IEEE1394, to a host and may have an attached external DRAM for compressed page storage. A bridge chip would have one or more ISI interfaces. The use of multiple ISI buses would allow the construction of 35 independent print systems within the one printer. The ISI-Bridge would be the ISIMaster for each of the ISI buses it interfaces to.

POSSIBLE SoPEC SYSTEMS

Several possible SoPEC based system architectures exist. The following sections outline some possible architectures. It is possible to have extra SoPEC devices in the system used for DRAM storage. The QA

40 chip configurations shown are indicative of the flexibility of LSS bus architecture, but not limited to those configurations.

A4 Simplex with 1 SoPEC device

In Figure 3, a single SoPEC device can be used to control two printhead ICs. The SoPEC receives compressed data through the USB device from the host. The compressed data is processed and transferred to the printhead.

5 A4 Duplex with 2 SoPEC devices

In Figure 4, two SoPEC devices are used to control two bi-lithic printheads, each with two printhead ICs. Each bi-lithic printhead prints to opposite sides of the same page to achieve duplex printing. The SoPEC connected to the USB is the ISIMaster SoPEC, the remaining SoPEC is an ISISlave. The ISIMaster receives all the compressed page data for both SoPECs and re-distributes the compressed data over the 10 Inter-SoPEC Interface (ISI) bus.

It may not be possible to print an A4 page every 2 seconds in this configuration since the USB 1.1 connection to the host may not have enough bandwidth. An alternative would be for each SoPEC to have its own USB 1.1 connection. This would allow a faster average print speed.

15 A3 Simplex with 2 SoPEC devices

15

In Figure 5, two SoPEC devices are used to control one A3 bi-lithic printhead. Each SoPEC controls only one printhead IC (the remaining PHI port typically remains idle). This system uses the SoPEC with the USB connection as the ISIMaster. In this dual SoPEC configuration the compressed page store data is split across 2 SoPECs giving a total of 4Mbyte page store, this allows the system to use compression 20 rates as in an A4 architecture, but with the increased page size of A3. The ISIMaster receives all the compressed page data for all SoPECs and re-distributes the compressed data over the Inter-SoPEC Interface (ISI) bus.

It may not be possible to print an A3 page every 2 seconds in this configuration since the USB 1.1 connection to the host will only have enough bandwidth to supply 2Mbytes every 2 seconds. Pages which

25

require more than 2MBytes every 2 seconds will therefore print more slowly. An alternative would be for each SoPEC to have its own USB 1.1 connection. This would allow a faster average print speed.

A3 Duplex with 4 SoPEC devices

30

In Figure 6 a 4 SoPEC system is shown. It contains 2 A3 bi-lithic printheads, one for each side of an A3 page. Each printhead contain 2 printhead ICs, each printhead IC is controlled by an independent SoPEC device, with the remaining PHI port typically unused. Again the SoPEC with USB 1.1 connection is the ISIMaster with the other SoPECs as ISISlaves. In total, the system contains 8Mbytes of compressed 35 page store (2Mbytes per SoPEC), so the increased page size does not degrade the system print quality, from that of an A4 simplex printer. The ISIMaster receives all the compressed page data for all SoPECs and re-distributes the compressed data over the Inter-SoPEC Interface (ISI) bus.

It may not be possible to print an A3 page every 2 seconds in this configuration since the USB 1.1 connection to the host will only have enough bandwidth to supply 2Mbytes every 2 seconds. Pages which require more than 2MBytes every 2 seconds will therefore print more slowly. An alternative would be for each SoPEC or set of SoPECs on the same side of the page to have their own USB 1.1 connection (as 40 ISISlaves may also have direct USB connections to the host). This would allow a faster average print speed.

SoPEC DRAM storage solution: A4 Simplex with 1 printing SoPEC and 1 memory SoPEC

Extra SoPECs can be used for DRAM storage e.g. in Figure 7 an A4 simplex printer can be built with a single extra SoPEC used for DRAM storage. The DRAM SoPEC can provide guaranteed bandwidth delivery of data to the printing SoPEC. SoPEC configurations can have multiple extra SoPECs used for DRAM storage.

5 ISI-Bridge chip solution: A3 Duplex system with 4 SoPEC devices

In Figure 8, an ISI-Bridge chip provides slave-only ISI connections to SoPEC devices. Figure 8 shows a ISI-Bridge chip with 2 separate ISI ports. The ISI-Bridge chip is the ISIMaster on each of the ISI busses it is connected to. All connected SoPECs are ISISlaves. The ISI-Bridge chip will typically have a high bandwidth connection to a host and may have an attached external DRAM for compressed page storage.

10 An alternative to having a ISI-Bridge chip would be for each SoPEC or each set of SoPECs on the same side of a page to have their own USB 1.1 connection. This would allow a faster average print speed.

Page Format and Printflow

When rendering a page, the RIP produces a page header and a number of bands (a non-blank page requires at least one band) for a page. The page header contains high level rendering parameters, and

15 each band contains compressed page data. The size of the band will depend on the memory available to the RIP, the speed of the RIP, and the amount of memory remaining in SoPEC while printing the previous band(s). Figure 9 shows the high level data structure of a number of pages with different numbers of bands in the page.

Each compressed band contains a mandatory band header, an optional bi-level plane, optional sets of

20 interleaved contone planes, and an optional tag data plane (for Netpage enabled applications). Since each of these planes is optional¹, the band header specifies which planes are included with the band.

Figure 10 gives a high-level breakdown of the contents of a page band.

A single SoPEC has maximum rendering restrictions as follows:

25

- 1 bi-level plane
- 1 contone interleaved plane set containing a maximum of 4 contone planes
- 1 tag data plane
- a bi-lithic printhead with a maximum of 2 printhead ICs

The requirement for single-sided A4 single SoPEC printing is

30

- average contone JPEG compression ratio of 10:1, with a local minimum compression ratio of 5:1 for a single line of interleaved JPEG blocks.
- average bi-level compression ratio of 10:1, with a local minimum compression ratio of 1:1 for a single line.

If the page contains rendering parameters that exceed these specifications, then the RIP or the Host PC 35 must split the page into a format that can be handled by a single SoPEC.

In the general case, the SoPEC CPU must analyze the page and band headers and generate an appropriate set of register write commands to configure the units in SoPEC for that page. The various bands are passed to the destination SoPEC(s) to locations in DRAM determined by the host.

¹Although a band must contain at least one plane

The host keeps a memory map for the DRAM, and ensures that as a band is passed to a SoPEC, it is stored in a suitable free area in DRAM. Each SoPEC is connected to the ISI bus or USB bus via its Serial communication Block (SCB). The SoPEC CPU configures the SCB to allow compressed data bands to pass from the USB or ISI through the SCB to SoPEC DRAM. Figure 11 shows an example data flow for a

5 page destined to be printed by a single SoPEC. Band usage information is generated by the individual SoPECs and passed back to the host.

SoPEC has an addressing mechanism that permits circular band memory allocation, thus facilitating easy memory management. However it is not strictly necessary that all bands be stored together. As long as 10 the appropriate registers in SoPEC are set up for each band, and a given band is contiguous², the memory can be allocated in any way.

SoPEC ASIC

The Small Office Home Office Print Engine Controller (SoPEC) is a page rendering engine ASIC that

15 takes compressed page images as input, and produces decompressed page images at up to 6 channels of bi-level dot data as output. The bi-level dot data is generated for the Memjet bi-lithic printhead. The dot generation process takes account of printhead construction, dead nozzles, and allows for fixative generation.

A single SoPEC can control 2 bi-lithic printheads and up to 6 color channels at 10,000 lines/sec³,

20 equating to 30 pages per minute. A single SoPEC can perform full-bleed printing of A3, A4 and Letter pages. The 6 channels of colored ink are the expected maximum in a consumer SOHO, or office Bi-lithic printing environment:

- CMY, for regular color printing.
- K, for black text, line graphics and gray-scale printing.
- 25 IR (infrared), for Netpage-enabled [5] applications.
- F (fixative), to enable printing at high speed. Because the bi-lithic printer is capable of printing so fast, a fixative may be required to enable the ink to dry before the page touches the page already printed. Otherwise the pages may bleed on each other. In low speed printing environments the fixative may not be required.

30 SoPEC is *color space agnostic*. Although it can accept contone data as CMYX or RGBX, where X is an optional 4th channel, it also can accept contone data in any print color space. Additionally, SoPEC provides a mechanism for arbitrary mapping of input channels to output channels, including combining dots for ink optimization, generation of channels based on any number of other channels etc. However, inputs are typically CMYK for contone input, K for the bi-level input, and the optional Netpage tag dots are typi-

35 cally rendered to an infra-red layer. A fixative channel is typically generated for fast printing applications. SoPEC is *resolution agnostic*. It merely provides a mapping between input resolutions and output resolutions by means of scale factors. The expected output resolution is 1600 dpi, but SoPEC actually has no knowledge of the physical resolution of the Bi-lithic printhead.

²Contiguous allocation also includes wrapping around in SoPEC's band store memory.

³10,000 lines per second equates to 30 A4/Letter pages per minute at 1600 dpi

SoPEC is *page-length agnostic*. Successive pages are typically split into bands and downloaded into the page store as each band of information is consumed and becomes free.

SoPEC provides an interface for synchronization with other SoPCs. This allows simple multi-SoPEC solutions for simultaneous A3/A4/Letter duplex printing. However, SoPEC is also capable of printing only

5 a portion of a page image. Combining synchronization functionality with partial page rendering allows multiple SoPCs to be readily combined for alternative printing requirements including simultaneous duplex printing and wide format printing.

PRINTING RATES

The required printing rate for SoPEC is 30 sheets per minute with an inter-sheet spacing of 4 cm. To

10 achieve a 30 sheets per minute print rate, this requires:

$$300\text{mm} \times 63 \text{ (dot/mm)} / 2 \text{ sec} = 105.8 \text{ microseconds per line, with no inter-sheet gap.}$$

$$340\text{mm} \times 63 \text{ (dot/mm)} / 2 \text{ sec} = 93.3 \text{ microseconds per line, with a 4 cm inter-sheet gap.}$$

A printline for an A4 page consists of 13824 nozzles across the page [2]. At a system clock rate of 160 MHz 13824 dots of data can be generated in 86.4 microseconds. Therefore data can be generated fast

15 enough to meet the printing speed requirement. It is necessary to deliver this print data to the print-heads. Printheads can be made up of 5:5, 6:4, 7:3 and 8:2 inch printhead combinations [2]. Print data is transferred to both print heads in a pair simultaneously. This means the longest time to print a line is determined by the time to transfer print data to the longest print segment. There are 9744 nozzles across a 7 inch printhead. The print data is transferred to the printhead at a rate of 106 MHz (2/3 of the system 20 clock rate) per color plane. This means that it will take 91.9 μ s to transfer a single line for a 7:3 printhead configuration. So we can meet the requirement of 30 sheets per minute printing with a 4 cm gap with a 7:3 printhead combination. There are 11160 across an 8 inch printhead. To transfer the data to the printhead at 106 MHz will take 105.3 μ s. So an 8:2 printhead combination printing with an inter-sheet gap will print slower than 30 sheets per minute.

25 SoPEC BASIC ARCHITECTURE

From the highest point of view the SoPEC device consists of 3 distinct subsystems

- CPU Subsystem
- DRAM Subsystem
- Print Engine Pipeline (PEP) Subsystem

30 See Figure 13 for a block level diagram of SoPEC.

CPU Subsystem

The CPU subsystem controls and configures all aspects of the other subsystems. It provides general support for interfacing and synchronising the external printer with the internal print engine. It also controls the low speed communication to the QA chips. The CPU subsystem contains various peripherals to aid the

35 CPU, such as GPIO (includes motor control), interrupt controller, LSS Master and general timers. The Serial Communications Block (SCB) on the CPU subsystem provides a full speed USB1.1 interface to the host as well as an Inter SoPEC Interface (ISI) to other SoPEC devices.

DRAM Subsystem

The DRAM subsystem accepts requests from the CPU, Serial Communications Block (SCB) and blocks 40 within the PEP subsystem. The DRAM subsystem (in particular the DIU) arbitrates the various requests

and determines which request should win access to the DRAM. The DIU arbitrates based on configured parameters, to allow sufficient access to DRAM for all requestors. The DIU also hides the implementation specifics of the DRAM such as page size, number of banks, refresh rates etc.

Print Engine Pipeline (PEP) subsystem

5 The Print Engine Pipeline (PEP) subsystem accepts compressed pages from DRAM and renders them to bi-level dots for a given print line destined for a printhead interface that communicates directly with up to 2 segments of a bi-lithic printhead.

The first stage of the page expansion pipeline is the CDU, LBD and TE. The CDU expands the JPEG-compressed contone (typically CMYK) layer, the LBD expands the compressed bi-level layer (typically K),

10 and the TE encodes Netpage tags for later rendering (typically in IR or K ink). The output from the first stage is a set of buffers: the CFU, SFU, and TFU. The CFU and SFU buffers are implemented in DRAM. The second stage is the HCU, which dithers the contone layer, and composites position tags and the bi-level spot0 layer over the resulting bi-level dithered layer. A number of options exist for the way in which compositing occurs. Up to 6 channels of bi-level data are produced from this stage. Note that not all 6

15 channels may be present on the printhead. For example, the printhead may be CMY only, with K pushed into the CMY channels and IR ignored. Alternatively, the position tags may be printed in K if IR ink is not available (or for testing purposes).

The third stage (DNC) compensates for dead nozzles in the printhead by color redundancy and error diffusing dead nozzle data into surrounding dots.

20 The resultant bi-level 6 channel dot-data (typically CMYK-IRF) is buffered and written out to a set of line buffers stored in DRAM via the DWU.

Finally, the dot-data is loaded back from DRAM, and passed to the printhead interface via a dot FIFO.

The dot FIFO accepts data from the LLU at the system clock rate (pc/k), while the PHI removes data from the FIFO and sends it to the printhead at a rate of 2/3 times the system clock rate.

25

SoPEC BLOCK DESCRIPTION

Looking at Figure 13, the various units are described here in summary form:

Table 3. Units within SoPEC

Subsystem	Unit Acronym	Unit Name	Description
DRAM	DIU	DRAM interface unit	Provides the interface for DRAM read and write access for the various SoPEC units, CPU and the SCB block. The DIU provides arbitration between competing units controls DRAM access.
	DRAM	Embedded DRAM	20Mbits of embedded DRAM,
CPU	CPU	Central Processing Unit	CPU for system configuration and control
	MMU	Memory Management Unit	Limits access to certain memory address areas in CPU user mode

RDU	Real-time Debug Unit	Facilitates the observation of the contents of most of the CPU addressable registers in SoPEC in addition to some pseudo-registers in realtime.	
TIM	General Timer	Contains watchdog and general system timers	
LSS	Low Speed Serial Interfaces	Low level controller for interfacing with the QA chips	
GPIO	General Purpose IOs	General IO controller, with built-in Motor control unit, LED pulse units and de-glitch circuitry	
ROM	Boot ROM	16 KBytes of System Boot ROM code	
ICU	Interrupt Controller Unit	General Purpose interrupt controller with configurable priority, and masking.	
CPR	Clock, Power and Reset block	Central Unit for controlling and generating the system clocks and resets and powerdown mechanisms	
PSS	Power Save Storage	Storage retained while system is powered down	
USB	Universal Serial Bus Device	USB device controller for interfacing with the host USB.	
ISI	Inter-SoPEC Interface	ISI controller for data and control communication with other SoPEC's in a multi-SoPEC system	
SCB	Serial Communication Block	Contains both the USB and ISI blocks.	
Print Engine Pipeline (PEP)	PCU	PEP controller	Provides external CPU with the means to read and write PEP Unit registers, and read and write DRAM in single 32-bit chunks.
	CDU	Contone decoder unit	Expands JPEG compressed contone layer and writes decompressed contone to DRAM
	CFU	Contone FIFO Unit	Provides line buffering between CDU and HCU
	LBD	Lossless Bi-level Decoder	Expands compressed bi-level layer.
	SFU	Spot FIFO Unit	Provides line buffering between LBD and HCU
	TE	Tag encoder	Encodes tag data into line of tag dots.
	TFU	Tag FIFO Unit	Provides tag data storage between TE and HCU
	HCU	Halftoner compositor unit	Dithers contone layer and composites the bi-level spot 0 and position tag dots.
	DNC	Dead Nozzle Compensator	Compensates for dead nozzles by color redundancy and error diffusing dead nozzle data into surrounding dots.
	DWU	Dotline Writer Unit	Writes out the 6 channels of dot data for a

		given printline to the line store DRAM
LLU	Line Loader Unit	Reads the expanded page image from line store, formatting the data appropriately for the bi-lithic printhead.
PHI	PrintHead Interface	Is responsible for sending dot data to the bi-lithic printheads and for providing line synchronization between multiple SoPECs. Also provides test interface to printhead such as temperature monitoring and Dead Nozzle Identification.

Clocking, Power and Reset (CPR)

The CPR block provides all of the clock, power enable and reset signals to the SoPEC device.

POWERDOWN MODES

5 The CPR block is capable of powering down certain sections of the SoPEC device. When a section is powered down (i.e. put in sleep mode) no state is retained(except the PSS storage), the CPU must re-initialize the section before it can be used again.

For the purpose of powerdown the SoPEC device is divided into sections:

Table 4. Powerdown sectioning

10

Section	Block
Print Engine Pipeline SubSystem (Section 0)	PCU
	CDU
	CFU
	LBD
	SFU
	TE
	TFU
	HCU
	DNC
	DWU
	LLU
	PHI
CPU-DRAM (Section 1)	DRAM
	CPU/MMU
	DIU
	TIM
	ROM
	LSS
	PSS

	ICU
ISI Subsystem (Section 2)	ISI (SCB)
	DMA Ctrl (SCB)
	GPIO
USB Subsystem (Section 3)	USB (SCB)

Note that the CPR block is not located in any section. All configuration registers in the CPR block are clocked by an ungatable clock and have special reset conditions.

Sleep mode

Each section can be put into sleep mode by setting the corresponding bit in the *SleepModeEnable*

5 register. To re-enable the section the sleep mode bit needs to be cleared and then the section should be reset by writing to the relevant bit in the *ResetSection* register. Each block within the section should then be re-configured by the CPU.

If the CPU system (section 1) is put into sleep mode, the SoPEC device will remain in sleep mode until a system level reset is initiated from the reset pin, or a wakeup reset by the SCB block as a result of activity

10 on either the USB or ISI bus. The watchdog timer cannot reset the device as it is in section 1 also, and will be in sleep mode.

If the CPU and ISI subsystem are in sleep mode only a reset from the USB or a hardware reset will re-activate the SoPEC device.

If all sections are put into sleep mode, then only a system level reset initiated by the reset pin will re-activate the SoPEC device.

Like all software resets in SoPEC the *ResetSection* register is active-low i.e. a 0 should be written to each bit position requiring a reset. The *ResetSection* register is self-resetting.

Sleep Mode powerdown procedure

When powering down a section, the section may retain its current state (although not guaranteed to). It is

20 possible when powering back up a section that inconsistencies between interface state machines could cause incorrect operation. In order to prevent such condition from happening, all blocks in a section must be disabled before powering down. This will ensure that blocks are restored in a benign state when powered back up.

In the case of PEP section units setting the Go bit to zero will disable the block. The DRAM subsystem

25 can be effectively disabled by setting the *RotationSync* bit to zero, and the SCB system disabled by setting the *DMAAccessEn* bits to zero turning off the DMA access to DRAM. Other CPU subsystem blocks without any DRAM access do not need to be disabled.

RESET SOURCE

The SoPEC device can be reset by a number of sources. When a reset from an internal source is initiated

30 the reset source register (*ResetSrc*) stores the reset source value. This register can then be used by the CPU to determine the type of boot sequence required.

CLOCK RELATIONSHIP

The crystal oscillator excites a 32MHz crystal through the *xtalin* and *xtalout* pins. The 32MHz output is used by the PLL to derive the master VCO frequency of 960MHz. The master clock is then divided to produce 320MHz clock (*clk320*), 160MHz clock (*clk160*) and 48MHz (*clk48*) clock sources.

5 The phase relationship of each clock from the PLL will be defined. The relationship of internal clocks *clk320*, *clk48* and *clk160* to *xtalin* will be undefined.

At the output of the clock block, the skew between each *pclk* domain (*pclk_section[2:0]* and *jclk*) should be within skew tolerances of their respective domains (defined as less than the hold time of a D-type flip flop).

10 The skew between *dclk* and *pclk* should also be less than the skew tolerances of their respective domains.

The *usbclk* is derived from the PLL output and has no relationship with the other clocks in the system and is considered asynchronous.

PLL CONTROL

15 The PLL in SoPEC can be adjusted by programming the *PLLRangeA*, *PLLRangeB*, *PLL_Tunebits* and *PLL_Mult* registers. If these registers are changed by the CPU the values are not updated until the *PLLUpdate* register is written to. Writing to the *PLLUpdate* register triggers the PLL control state machine to update the PLL configuration in a safe way. When an update is active (as indicated by *PLLUpdate* register) the CPU must not change any of the configuration registers, doing so could cause the PLL to 20 lose lock indefinitely, requiring a hardware reset to recover. Configuring the PLL registers in an inconsistent way can also cause the PLL to lose lock, care must be taken to keep the PLL configuration within specified parameters.

The VCO frequency of the PLL is calculated by the number of divider in the feedback path. PLL output A is used as the feedback source.

25 $VCO_{freq} = REFCLK \times PLL_Mult \times PLLRangeA \times \text{External divider}$

$$VCO_{freq} = 32 \times 3 \times 10 \times 1 = 960 \text{ Mhz.}$$

In the default PLL setup, *PLL_Mult* is set to 3, *PLLRangeA* is set to 3 which corresponds to a divide by 10, *PLLRangeB* is set to 5 which corresponds to a divide by 3.

$$PLL_{outA} = VCO_{freq} / PLLRangeA = 960 \text{ Mhz} / 10 = 96 \text{ Mhz}$$

30 $PLL_{outB} = VCO_{freq} / PLLRangeB = 960 \text{ Mhz} / 3 = 320 \text{ Mhz}$

See [16] for complete PLL setup parameters.

IMPLEMENTATION

Definitions of I/O

35 Table 5. CPR I/O definition

Port name	Pins	I/O	Description
Clocks and Resets			
XtalIn	1	In	Crystal input, direct from IO pin.
XtalOut	1	Inout	Crystal output, direct to IO pin.
<i>pclk_section[3:0]</i>	4	Out	System clocks for each section

Doclk	1	Out	Data out clock (2x <i>pc/k</i>) for the PHI block
Jclk	1	Out	Gated version of system clock used to clock the JPEG decoder core in the CDU
Usbclk	1	Out	USB clock, nominally at 48 Mhz
jclk_enable	1	In	Gating signal for <i>jclk</i> . When 1 <i>jclk</i> is enabled
reset_n	1	In	Reset signal from the <i>reset_n</i> pin
usb_cpr_reset_n	1	In	Reset signal from the USB block
isi_cpr_reset_n	1	In	Reset signal from the ISI block
tim_cpr_reset_n	1	In	Reset signal from watch dog timer.
gpio_cpr_wakeup	1	In	SoPEC wake up from the GPIO, active high.
prst_n_section[3:0]	4	Out	System resets for each section, synchronous active low
dorst_n	1	Out	Reset for PHI block, synchronous to <i>doclk</i>
jrst_n	1	Out	Reset for JPEG decoder core in CDU block, synchronous to <i>jclk</i>
usbrst_n	1	Out	Reset for the USB block, synchronous to <i>usbclk</i>
CPU interface			
cpu_adr[5:2]	3	In	CPU address bus. Only 4 bits are required to decode the address space for the CPR block
cpu_dataout[31:0]	32	In	Shared write data bus from the CPU
cpr_cpu_data[31:0]	32	Out	Read data bus to the CPU
cpu_rwn	1	In	Common read/not-write signal from the CPU
cpu_cpr_sel	1	In	Block select from the CPU. When <i>cpr_cpr_sel</i> is high both <i>cpu_adr</i> and <i>cpu_dataout</i> are valid
cpr_cpu_rdy	1	Out	Ready signal to the CPU. When <i>cpr_cpu_rdy</i> is high it indicates the last cycle of the access. For a write cycle this means <i>cpr_cpu_data</i> has been registered by the block and for a read cycle this means the data on <i>cpr_cpu_data</i> is valid.
cpr_cpu_berr	1	Out	Bus error signal to the CPU indicating an invalid access.
cpu_acode[1:0]	2	In	CPU Access Code signals. These decode as follows: 00 - User program access 01 - User data access 10 - Supervisor program access 11 - Supervisor data access
cpr_cpu_debug_valid	1	Out	Debug Data valid on <i>cpr_cpu_data</i> bus. Active high

Configuration registers

The configuration registers in the CPR are programmed via the CPU interface. Note that since addresses in SoPEC are byte aligned and the CPU only supports 32-bit register reads and writes, the lower 2 bits of the CPU address bus are not required to decode the address space for the CPR. When reading a register

5 that is less than 32 bits wide zeros should be returned on the upper unused bit(s) of *cpr_pcu_data*. Table 6 lists the configuration registers in the CPR block.

The CPR block will only allow supervisor data mode accesses (i.e. *cpu_acode[1:0] = SUPERVISOR_DATA*). All other accesses will result in *cpr_cpu_berr* being asserted .

Table 6. CPR Register Map

10

Address CPR_base +	Register	#bits	Reset	Description
0x00	SleepModeEnable	4	0x0 ^a	Sleep Mode enable, when high a section of logic is put into powerdown. Bit 0 - Controls section 0 Bit 1 - Controls section 1 Bit 2 - Controls section 2 Bit 3 - Controls section 3 Note that the SleepModeEnable register has special reset conditions. See Section 16.5.6 for details
0x04	ResetSrc	5	0x1 ^a	Reset Source register, indicating the source of the last reset (or wake-up) Bit 0 - External Reset Bit 1 - USB wakeup reset Bit 2 - ISI wakeup reset Bit 3 - Watchdog timer reset Bit 4 - GPIO wake-up (Read Only Register)
0x08	ResetSection	4	0xF	Active-low synchronous reset for each section, self-resetting. Bit 0 - Controls section 0 Bit 1 - Controls section 1 Bit 2 - Controls section 2 Bit 3 - Controls section 3
0x0C	DebugSelect[5:2]	4	0x0	Debug address select. Indicates the address of the register to report on the <i>cpr_pcu_data</i> bus when it is not otherwise being used.
PLL Control				
0x10	PLL Tune Bits	10	0x3BC	PLL tuning bits

0x14	PLLRangeA	4	0x3	PLLOUT A frequency selector (defaults to 60Mhz to 125Mhz)
0x18	PLLRangeB	3	0x5	PLLOUT B frequency selector (defaults to 200Mhz to 400Mhz)
0x1C	PLLMultiplier	5	0x03	PLL multiplier selector, defaults to $refclk \times 3$
0x20	PLLUpdate	1	0x0	<p>PLL update control. A write (of any value) to this register will cause the PLL to lose lock for ~100us. Reading the register indicates the status of the update.</p> <p>0 - PLL update complete 1 - PLL update active No writes to <i>PLL Tune Bits</i>, <i>PLL Range A</i>, <i>PLL Range B</i>, <i>PLLMultiplier</i> or <i>PLLUpdate</i> are allowed while the PLL update is active.</p>

a. Reset value depends on reset source. External reset shown.

CPR Sub-block partition

reset_n deglitch

The external *reset_n* signal is deglitched for about 1 μ s. *reset_n* must maintain a state for 1us second before the state is passed into the rest of the device. All deglitch logic is clocked on *bufrefclk*.

Sync reset

The reset synchronizer retimes an asynchronous reset signal to the clock domain that it resets. The circuit prevents the inactive edge of reset occurring when the clock is rising

Reset generator logic

10 The reset generator logic is used to determine which clock domains should be reset, based on configured reset values (*reset_section_n*), the external reset (*reset_n*), watchdog timer reset (*tim_cpr_reset_n*), the USB reset (*usb_cpr_reset_n*), the GPIO wakeup control (*gpio_cpr_wakeup*) and the ISI reset (*isi_cpr_reset_n*). The reset direct from the IO pin (*reset_n*) is synchronized and de-glitched before feeding the reset logic.

15 All resets are lengthened to at least 16 *pclk* cycles, regardless of the duration of the input reset. The clock for a particular section must be running for the reset to have an effect. The clocks to each section can be enabled/disabled using the *SleepModeEnable* register.

Resets from the ISI or USB block reset everything except its own section (section 2 or 3).

Table 7. Reset domains

20

Reset signal	Domain
<i>reset_dom[0]</i>	Section 0 <i>pclk</i> domain (PEP)
<i>reset_dom[1]</i>	Section 1 <i>pclk</i> domain (CPU)

reset_dom[2]	Section 2 pclk domain (ISI)
reset_dom[3]	Section 3 usbcclk/pclk domain (USB)
reset_dom[4]	doclk domain
reset_dom[5]	jclk domain

The logic is given by

```

5      if (reset_dg_n == 0) then
          reset_dom[5:0]      = 0x00      // reset everything
          reset_src[4:0]       = 0x01
          cfg_reset_n          = 0
          sleep_mode_en[3:0]   = 0x0        // re-awaken all sections
      elsif (tim_cpr_reset_n == 0) then
          reset_dom[5:0]      = 0x00      // reset everything except CPR
10     config
          reset_src[4:0]       = 0x08
          cfg_reset_n          = 1        // CPR config stays the same
          sleep_mode_en[1]     = 0        // re-awaken section 1 only
          (awake already)
15     elsif (usb_cpr_reset_n == 0) then
          reset_dom[5:0]      = 0x08      // all except USB domain + CPR
          config
          reset_src[4:0]       = 0x02
          cfg_reset_n          = 1        // CPR config stays the same
20     sleep_mode_en[1]     = 0        // re-awaken section 1 only,
          section 3 is awake
          elsif (isi_cpr_reset_n == 0) then
              reset_dom[5:0]    = 0x04      // all except ISI domain + CPR
              config
              reset_src[4:0]     = 0x04
              cfg_reset_n        = 1        // CPR config stays the same
              sleep_mode_en[1]   = 0        // re-awaken section 1 only,
25     section 2 is awake
              elsif (gpio_cpr_wakeup = 1) then
                  reset_dom[5:0]    = 0x3C      // PEP and CPU sections only
                  reset_src[4:0]     = 0x10
                  cfg_reset_n        = 1        // CPR config stays the same
                  sleep_mode_en[1]   = 0        // re-awaken section 1 only,
30     section 2 is awake
                  else
                      // propagate resets from reset section register
35     reset_dom[5:0]      = 0x3F      // default to on

```

```

    cfg_reset_n          = 1           // CPR cfg registers
    are not in any section
    sleep_mode_en[3:0]   = sleep_mode_en[3:0] // stay the same by
    default
5      if (reset_section_n[0] == 0) then
        reset_dom[5] = 0           // jclk domain
        reset_dom[4] = 0           // doclk domain
        reset_dom[0] = 0           // pclk section 0 domain
    if (reset_section_n[1] == 0) then
        reset_dom[1] = 0           // pclk section 1 domain
10     if (reset_section_n[2] == 0) then
        reset_dom[2] = 0           // pclk section 2 domain (ISI)
    if (reset_section_n[3] == 0) then
        reset_dom[3] = 0           // USB domain
15

```

Sleep logic

The sleep logic is used to generate gating signals for each of SoPECs clock domains. The gate enable (*gate_dom*) is generated based on the configured *sleep_mode_en* and the internally generated *jclk_enable* signal.

20 The logic is given by

```

    // clock gating for sleep modes
    gate_dom[5:0] = 0x0           // default to all clocks on
    if (sleep_mode_en[0] == 1) then // section 0 sleep
        gate_dom[0] = 1           // pclk section 0
25      gate_dom[4] = 1           // doclk domain
        gate_dom[5] = 1           // jclk domain
    if (sleep_mode_en[1] == 1) then // section 1 sleep
        gate_dom[1] = 1           // pclk section 1
    if (sleep_mode_en[2] == 1) then // section 2 sleep
        gate_dom[2] = 1           // pclk section 2
30      if (sleep_mode_en[3] == 1) then // section 3 sleep
        gate_dom[3] = 1           // usb section 3
    // the jclk can be turned off by CDU signal
    if (jclk_enable == 0) then
35      gate_dom[5] = 1

```

The clock gating and sleep logic is clocked with the *master_pclk* clock which is not gated by this logic, but is synchronous to other *pclk_section* and *jclk* domains.

Once a section is in sleep mode it cannot generate a reset to restart the device. For example if section 1 is in sleep mode then the watchdog timer is effectively disabled and cannot trigger a reset.

40 Clock gate logic

The clock gate logic is used to safely gate clocks without generating any glitches on the gated clock. When the enable is high the clock is active otherwise the clock is gated.

Clock generator Logic

The clock generator block contains the PLL, crystal oscillator, clock dividers and associated control logic. The PLL VCO frequency is at 960MHz locked to a 32 MHz *refclk* generated by the crystal oscillator. In test mode the *xtalin* signal can be driven directly by the test clock generator, the test clock will be reflected on the *refclk* signal to the PLL.

5 *Clock divider A*

The clock divider A block generates the 48MHz clock from the input 96MHz clock (*pllouta*) generated by the PLL. The divider is enabled only when the PLL has acquired lock.

Clock divider B

The clock divider B block generates the 160MHz clocks from the input 320MHz clock (*plloutb*) generated by the PLL. The divider is enabled only when the PLL has acquired lock.

PLL control state machine

The PLL will go out of lock whenever *pll_reset* goes high (the PLL reset is the only active high reset in the device) or if the configuration bits *pll_rangea*, *pll_rangeb*, *pll_mult*, *pll_tune* are changed. The PLL control state machine ensures that the rest of the device is protected from glitching clocks while the PLL is being reset or its configuration is being changed.

In the case of a hardware reset (the reset is deglitched), the state machine first disables the output clocks (via the *clk_gate* signal), it then holds the PLL in reset while its configuration bits are reset to default values. The state machine then releases the PLL reset and waits approx. 100us to allow the PLL to regain lock. Once the lock time has elapsed the state machine re-enables the output clocks and resets the remainder of the device via the *reset_dg_n* signal.

When the CPU changes any of the configuration registers it must write to the PLLUpdate register to allow the state machine to update the PLL to the new configuration setup. If a PLLUpdate is detected the state machine first gates the output clocks. It then holds the PLL in reset while the PLL configuration registers are updated. Once updated the PLL reset is released and the state machine waits approx 100us for the

25 PLL to regain lock before re-enabling the output clocks. Any write to the PLLUpdate register will cause the state machine to perform the update operation regardless of whether the configuration values changed or not.

All logic in the clock generator is clocked on *bufrefclk* which is always an active clock regardless of the state of the PLL.

30

SECURITY OVERVIEW

Introduction

35 A number of hardware, software and protocol solutions to security issues have been developed. These range from authorization and encryption protocols for enabling secure communication between hardware and software modules, to physical and electrical systems that protect the integrity of integrated circuits and other hardware.

40 It should be understood that in many cases, principles described with reference to hardware such as integrated circuits (ie, chips) can be implemented wholly or partly in software running on, for example, a computer. Mixed systems in which software and hardware (and combinations) embody various entities, modules and units can also be constructed using many of these principles, particularly in relation to

authorization and authentication protocols. The particular extent to which the principles described below can be translated to or from hardware or software will be apparent to one skilled in the art, and so will not always explicitly be explained.

5 It should also be understood that many of the techniques disclosed below have application to many fields other than printing. Some specific examples are described towards the end of this description.

A "QA Chip" is a quality assurance chip can allows certain security functions and protocols to be implemented.

10

QA CHIP TERMINOLOGY

The Authentication Protocols documents [5] and [6] refer to QA Chips by their function in particular protocols:

- For authenticated reads in [5], ChipR is the QA Chip being read from, and ChipT is the QA Chip that identifies whether the data read from ChipR can be trusted. ChipR and ChipT are referred to as Untrusted QA Device and Trusted QA Device respectively in [6].
- For replacement of keys in [5], ChipP is the QA Chip being programmed with the new key, and ChipF is the factory QA Chip that generates the message to program the new key. ChipF is referred to as the Key Programmer QA Device in [6].
- For upgrades of data in memory vectors in [5], ChipU is the QA Chip being upgraded, and ChipS is the QA Chip that signs the upgrade value. ChipS is referred to as the Value Upgrader QA Device and Parameter Upgrader QA Device in [6].

Any given physical QA Chip will contain functionality that allows it to operate as an entity in some number of these protocols.

Therefore, wherever the terms ChipR, ChipT, ChipP, ChipF, ChipU and ChipS are used in this document, 30 they are referring to *logical* entities involved in an authentication protocol as defined in [5] and [6].

Physical QA Chips are referred to by their location. For example, each ink cartridge may contain a QA Chip referred to as an INK_QA, with all INK_QA chips being on the same physical bus. In the same way, the QA Chip inside the printer is referred to as PRINTER_QA, and will be on a separate bus to the 35 INK_QA chips.

Background

The QA Chip can be used to hold authentication keys together with circuitry specially designed to prevent 40 copying. The chip is manufactured using a standard Flash memory manufacturing process, and is low cost enough to be included in consumables such as ink and toner cartridges. The implementation is

approximately 1mm² in a 0.25 micron flash process, and has an expected die manufacturing cost of approximately 10 cents in 2003.

Once programmed, the QA Chips as described here are compliant with the NSA export guidelines since they do not constitute a strong encryption device. They can therefore be practically manufactured in the

5 USA (and exported) or anywhere else in the world.

Note that although the QA Chip is designed for use in authentication systems, it is microcoded, and can therefore be programmed for a variety of applications.

Nomenclature

The following symbolic nomenclature is used throughout this document:

10 Table 8. Summary of symbolic nomenclature

Symbol	Description
$F[X]$	Function F, taking a single parameter X
$F[X, Y]$	Function F, taking two parameters, X and Y
$X Y$	X concatenated with Y
$X \wedge Y$	Bitwise X AND Y
$X \vee Y$	Bitwise X OR Y (inclusive-OR)
$X \oplus Y$	Bitwise X XOR Y (exclusive-OR)
$\neg X$	Bitwise NOT X (complement)
$X \leftarrow Y$	X is assigned the value Y
$X \leftarrow \{Y, Z\}$	The domain of assignment inputs to X is Y and Z
$X = Y$	X is equal to Y
$X \neq Y$	X is not equal to Y
$\downarrow X$	Decrement X by 1 (floor 0)
$\uparrow X$	Increment X by 1 (modulo register length)
Erase X	Erase Flash memory register X
SetBits[X, Y]	Set the bits of the Flash memory register X based on Y
$Z \leftarrow \text{ShiftRight}[X, Y]$	Shift register X right one bit position, taking input bit Y and placing the output bit in Z

LOGICAL INTERFACE

The QA Chip has a physical and a logical external interface. The physical interface defines how the QA

15 Chip can be connected to a physical System, while the logical interface determines how that System can communicate with the QA Chip. This section deals with the logical interface.

OPERATING MODES

The QA Chip has four operating modes - *Idle Mode*, *Program Mode*, *Trim Mode* and *Active Mode*.

- *Active Mode* is entered on power-on Reset when the fuse has been blown, and whenever a

20 specific authentication command arrives from the System. Program code is only executed in *Active Mode*. When the reset program code has finished, or the results of the command have been returned to the System, the chip enters *Idle Mode* to wait for the next instruction.

- *Idle Mode* is used to allow the chip to wait for the next instruction from the System.
- *Trim Mode* is used to determine the clock speed of the chip and to trim the frequency during the initial programming stage of the chip (when Flash memory is garbage). The clock frequency *must* be trimmed via Trim Mode *before* Program Mode is used to store the program code.

5 • *Program Mode* is used to load up the operating program code, and is required because the operating program code is stored in Flash memory instead of ROM (for security reasons).

Apart from while the QA Chip is executing Reset program code, it is always possible to interrupt the QA Chip and change from one mode to another.

Active Mode

10 Active Mode is entered in any of the following three situations:

- power-on Reset when the fuse has been blown
- receiving a command consisting of a global id write byte (0x00) followed by the ActiveMode command byte (0x06)
- receiving a command consisting of a local id byte write followed by some number of bytes representing opcode and data.

In all cases, Active Mode causes execution of program code previously stored in the flash memory via Program Mode.

If Active Mode is entered by power-on Reset or the global id mechanism, the QA Chip executes specific reset startup code, typically setting up the local id and other IO specific data. The reset startup code

20 cannot be interrupted except by a power-down condition. The power-on reset startup mechanism cannot be used before the fuse has been blown since the QA Chip cannot tell whether the flash memory is valid or not. In this case the globalid mechanism must be used instead.

If Active Mode is entered by the local id mechanism, the QA Chip executes specific code depending on the following bytes, which function as opcode plus data. The interpretation of the following bytes depends

25 on whatever software happens to be stored in the QA Chip.

Idle Mode

The QA Chip starts up in *Idle Mode* when the fuse has not yet been blown, and returns to *Idle Mode* after the completion of another mode. When the QA Chip is in *Idle Mode*, it waits for a command from the master by watching the low speed serial line for an id that matches either the global id (0x00), or the

30 chip's local id.

- If the primary id matches the global id (0x00, common to all QA Chips), and the following byte from the master is the Trim Mode id byte, and the fuse has not yet been blown, the QA Chip enters *Trim Mode* and starts counting the number of internal clock cycles until the next byte is received. Trim

35 Mode cannot be entered if the fuse has been blown.

- If the primary id matches the global id (0x00, common to all QA Chips), and the following byte from the master is the Program Mode id byte, and the fuse has not yet been blown, the QA Chip enters *Program Mode*. Program Mode cannot be entered if the fuse has been blown.

40 • If the primary id matches the global id (0x00, common to all QA Chips), and the following byte from the master is the Active Mode

id bytes, the QA Chip enters *Active Mode* and executes startup code, allowing the chip to set itself into a state to subsequently receive authentication commands (includes setting a local id and a trim value).

- If the primary id matches the chip's local id, the QA Chip enters *Active Mode*, allowing the subsequent command to be executed.

5

The valid 8-bit serial mode values sent after a global id are as shown in Table 9:

Table 9. Command byte values to place chip in specific mode

Value	Interpretation
10101011 (0xAB)	Trim Mode (only functions when the fuse has not been blown)
10001101 (0xAD)	Program Mode (only functions when the fuse has not been blown)
00000110 (0x06)	Active Mode (resets the chip & loads the localId)

10 Trim Mode

Trim Mode is enabled by sending a global id byte (0x00) followed by the Trim Mode command byte (0xAB). Trim Mode can only be entered while the fuse has not yet been blown.

The purpose of Trim Mode is to set the trim value (an internal register setting) of the internal ring oscillator so that Flash erasures and writes are of the correct duration. This is necessary due to the 2:1 variation of the clock speed due to process variations. If writes or erasures are too long, the Flash memory will wear out faster than desired, and in some cases can even be damaged. Note that the 2:1 variation due to temperature still remains, so the effective operating speed of the chip is 7-14 MHz around a nominal 10MHz.

Trim Mode works by measuring the number of system clock cycles that occur inside the chip from the

20 receipt of the Trim Mode command byte until the receipt of a data byte. When the data byte is received, the data byte is copied to the trim register and the current value of the count is transmitted to the outside world.

Once the count has been transmitted, the QA Chip returns to *Idle Mode*.

At reset, the internal trim register setting is set to a known value r . The external user can now perform the 25 following operations:

- send the global id+write followed by the Trim Mode command byte
- send the 8-bit value v over a specified time t
- send a stop bit to signify no more data
- send the global id+read followed by the Trim Mode command byte
- receive the count c
- send a stop bit to signify no more data

30

At the end of this procedure, the trim register will be v , and the external user will know the relationship between external time t and internal time c . Therefore a new value for v can be calculated.

The Trim Mode procedure can be repeated a number of times, varying both t and v in known ways, measuring the resultant c . At the end of the process, the final value for v is established (and stored in the trim register for subsequent use in Program Mode). This value v must also be written to the flash for later use (every time the chip is placed in Active Mode for the first time after power-up).

5 Program Mode

Program Mode is enabled by sending a global id byte (0x00) followed by the Program Mode command byte.

If the QA Chip knows already that the fuse has been blown, it simply does not enter Program Mode. If the QA Chip does not know the state of the fuse, it determines whether or not the internal fuse has been

blown by reading 32-bit word 0 of the information block of flash memory. If the fuse has been blown the remainder of data from the Program Mode command is ignored, and the QA Chip returns to *Idle Mode*. If the fuse is still intact, the chip enters Program Mode and erases the entire contents of Flash memory. The QA Chip then validates the erasure. If the erasure was successful, the QA Chip receives up to 4096 bytes of data corresponding to the new program code and variable data. The bytes are transferred in order $byte_0$ to $byte_{4095}$.

Once all bytes of data have been loaded into Flash, the QA Chip returns to *Idle Mode*.

Note that Trim Mode functionality must be performed before a chip enters Program Mode for the first time. Otherwise the erasure and write durations could be incorrect.

Once the desired number of bytes have been downloaded in Program Mode, the LSS Master must wait

for 80 μ s (the time taken to write two bytes to flash at nybble rates) before sending the new transaction (e.g. Active Mode). Otherwise the last nybbles may not be written to flash.

After Manufacture

Directly after manufacture the flash memory will be invalid and the fuse will not have been blown.

Therefore power-on-reset will not cause Active Mode. Trim Mode must therefore be entered first, and only after a suitable trim value is found, should Program Mode be entered to store a program. Active Mode can be entered if the program is known to be valid.

LOGICAL VIEW OF CPU

The QA Chip is a 32-bit microprocessor with on-board RAM for scratch storage, on-board flash for program storage, a serial interface, and specific security enhancements.

The high level commands that a user of an QA Chip sees are all implemented as small programs written in the CPU instruction set.

The following sections describe the memory model, the various registers, and the instruction set of the CPU.

Memory Model

The QA Chip has its own internal memory, broken into the following conceptual regions:

- *RAM variables* (3Kbits = 96 entries at 32-bits wide), used for scratch storage (e.g. HMAC-SHA1 processing).
- *Flash memory* (8Kbytes main block + 128 bytes info block) used to hold the non-volatile authentication variables (including program keys etc), and program code. Only 4 KBytes + 64 bytes is visible to the program addressing space due to shadowing. Shadowing is where half of each byte is used to validate and verify the other half, thus protecting against certain forms of physical and

logical attacks. As a result, two bytes are read to obtain a single byte of data (this happens transparently).

7.1 RAM

The RAM region consists of 96×32 -bit words required for the general functioning of the QA Chip, *but*

5 *only during the operation of the chip*. RAM is volatile memory: once power is removed, the values are lost. Note that in actual fact memory retains its value for some period of time after power-down, but cannot be considered to be available upon power-up. This has issues for security that are addressed in other sections of this document.

RAM is typically used for temporary storage of variables during chip operation. Short programs can also

10 be stored and executed from the RAM.

RAM is addressed from 0 to 5F. Since RAM is in an unknown state upon a RESET (RstL), program code should not assume the contents to be 0. Program code can, however, set the RAM to be a particular known state during execution of the reset command (guaranteed to be received before any other commands).

15 FLASH VARIABLES

The flash memory region contains the non-volatile information in the QA Chip. Flash memory retains its value after a RESET or if power is removed, and can be expected to be unchanged when the power is next turned on.

Byte 0 of main memory is the first byte of the program run for the command dispatcher. Note that the

20 command dispatcher is always run with shadows enabled.

Bytes 0-7 of the information block flash memory is reserved as follows:

- byte 0-3 = fuse. A value of 0x5555AAAA indicates that the fuse has been blown (think of a physical fuse whose wire is no longer intact).
- bytes 4-7 = random number used to XOR all data for RAM and flash memory accesses

After power-on reset (when the fuse is blown) or upon receipt of a globalId Active command, the 32-bit data from bytes 4-7 in the information block of Flash memory is loaded into an internal ChipMask register.

In Active Mode (the chip is executing program code), all data read from the flash and RAM is XORed with the ChipMask register, and all data written to the flash and RAM is XORed with the ChipMask register before being written out. This XORing happens completely transparently to the program code. Main flash memory byte 0 onward is the start of program code. Note that byte 0 onward needs to be valid after being XORed with the appropriate bytes of ChipMask.

Even though CPU access is in 8-bit and 32-bit quantities, the data is actually stored in flash a nybble-at-a-

35 time. Each nybble write is written as a byte containing 4 sets of b/¬b pairs. Thus every byte write to flash is writing a nybble to real and shadow. A write mask allows the individual targetting of nybble-at-a-time writes.

The checking of flash vs shadow flash is automatically carried out each read (each byte contains both flash and shadow flash). If all 8 bits are 1, the byte is considered to be in its erased form⁴, and returns 0 as

⁴TSMC's flash memory has an erased state of all 1s

the nybble. Otherwise, the value returned for the nybble depends on the size of the overall access and the setting of bit 0 of the 8-bit WriteMask.

- All 8-bit accesses (i.e. instruction and program code fetches) are checked to ensure that each byte read from flash is 4 sets of $b/\neg b$ pairs. If the data is not of this form, the chip hangs until a new command is issued over the serial interface.
- With 32-bit accesses (i.e. data used by program code), each byte read from flash is checked to ensure that it is 4 sets of $b/\neg b$ pairs. A setting of $\text{WriteMask}_0 = 0$ means that if the data is not valid, then the chip will hang until a new command is issued over the serial interface. A setting of $\text{WriteMask}_0 = 1$ means that each invalid nybble is replaced by the upper nybble of the WriteMask. This allows recovery after a write or erasure is interrupted by a power-down.

Registers

A number of registers are defined for use by the CPU. They are used for control, temporary storage,

15 arithmetic functions, counting and indexing, and for I/O.

These registers do not need to be kept in non-volatile (Flash) memory. They can be read or written without the need for an erase cycle (unlike Flash memory). Temporary storage registers that contain secret information still need to be protected from physical attack by Tamper Prevention and Detection circuitry and parity checks.

20 All registers are cleared to 0 on a RESET. However, program code should not assume any RAM contents have any particular state, and should set up register values appropriately. In particular, at the startup entry point, the various address registers need to be set up from unknown states.

GO

A 1-bit GO register is 1 when the program is executing, and 0 when it is not. Programs can clear the GO 25 register to halt execution of program code once the command has finished executing.

ACCUMULATOR AND Z FLAG

The Accumulator is a 32-bit general-purpose register that can be thought of as the single data register. It is used as one of the inputs to all arithmetic operations, and is the register used for transferring information between memory registers.

30 The Z register is a 1-bit flag, and is updated each time the Accumulator is written to. The Z register contains the zero-ness of the Accumulator. Z = 1 if the last value written to the Accumulator was 0, and 0 if the last value written was non-0.

Both the Accumulator and Z registers are directly accessible from the instruction set.

ADDRESS REGISTERS

35 Program Counter Array and Stack Pointer

A 12-level deep 12-bit Program Counter Array (PCA) is defined. It is indexed by a 4-bit Stack Pointer (SP). The current Program Counter (PC), containing the address of the currently executing instruction, is effectively PCA[SP]. A single register bit, PCRamSel determines whether the program is executing from flash or RAM (0 = flash, 1 = RAM).

The PC is affected by calling subroutines or returning from them, and by executing branching instructions.

The SP is affected by calling subroutines or returning from them. There is no bounds checking on calling too many subroutines: the oldest entry in the execution stack will be lost.

The entry point for program code is defined to be address 0 in Flash. This entry point is used whenever

5 the master signals a new transaction.

A0-A3

There are 4 8-bit address registers. Each register has an associated memory mode bit designating the address as in Flash (0) or RAM (1).

10 When an An register is pointing to an address in RAM, it holds the word number. When it is pointing to an address in Flash, it points to a set of 32-bit words that start at a 128-bit (16 byte) alignment.

The A0 register has a special use of direct offset e.g. access is possible to (A0),0-7 which is the 32-bit word pointed to by A0 offset by the specified number of words.

WriteMask

15 The WriteMask register is used to determine how many nybbles will be written during a 32-bit write to Flash, and whether or not an invalid nibble will be replaced during a read from Flash.

During writes to flash, bit n (of 8) determines whether nibble n is written. The unit of writing is a nibble since half of each byte is used for shadow data. A setting of 0xFF means that all 32-bits will be written to flash (as 8 sets of nibble writes).

20 During 32-bit reads from flash (occurs as 8 reads), the value of WriteMask₀ is used to determine whether a read of invalid data is replaced by the upper nibble of WriteMask. If 0, a read of invalid data is not replaced, and the chip hangs until a new command is issued over the serial interface. If 1, a read of invalid data is replaced by the upper nibble of the WriteMask.

Thus a WriteMask setting of 0 (reset setting) means that no writes will occur to flash, and all reads are not replaced (causing the program to hang if an invalid value is encountered).

COUNTERS

A number of special purpose counters/index registers are defined:

Table 10. Counter/Index registers

30

Name	Register Size	Bits	Description
C1	1×3	3	Counter used to index arrays and general purpose counter
C2	1×6	6	General purpose counter and can be used to index arrays

All these counter registers are directly accessible from the instruction set. Special instructions exist to load them with specific values, and other instructions exist to decrement or increment them, or to branch depending on the whether or not the specific counter is zero.

There are also 2 special flags (not registers) associated with c1 and c2, and these flags hold the zero-ness of c1 or c2. The flags are used for loop control, and are listed here, for although they are not registers, they can be tested like registers.

Table 11. Flags for testing C1 and C2

5

Name	Description
C1Z	1 = C1 is current zero, 0 = C1 is currently non-zero.
C2Z	1 = C2 is current zero, 0 = C2 is currently non-zero.

RTMP

The single bit register RTMP allows the implementation of LFSRs and multiple precision shift registers.

During a rotate right (ROR) instruction with operand of RB, the bit shifted out (formally bit 0) is written to the

10 RTMP register. The bit currently in the RTMP register becomes the new bit 31 of the Accumulator.

Performing multiple ROR RB commands over several 32-bit values implements a multiple precision rotate/shift right.

The XRB operand operates in the same way as RB, in that the current value in the RTMP register becomes the new bit 31 of the Accumulator. However with the XRB instruction, the bit formally known as bit 0 does not 15 simply replace RTMP (as in the RB instruction). Instead, it is XORed with RTMP, and the result stored in RTMP, thereby allowing the implementation of long LFSRs.

REGISTERS USED FOR I/O

Several registers are defined for communication between the master and the QA Chip. These registers are LocalId, InByte and OutByte.

20 LocalId (7 bits) defines the chip-specific id that this particular QA Chip will accept commands for. InByte (8 bits) provides the means for the QA Chip to obtain the next byte from the master. OutByte (8 bits) provides the means for the QA Chip to send a byte of data to the master.

From the QA Chip's point of view:

- Reads from InByte will hang until there is 1 byte of data present from the master.
- Writes to OutByte will hang if the master has not already consumed the last OutByte.

25 When the master begins a new command transaction, any existing data in InByte and OutByte is lost, and the PC is reset to the entry point in the code, thus ensuring correct framing of data.

REGISTERS USED FOR TRIMMING CLOCK SPEED

30 A single 8-bit Trim register is used to trim the ring oscillator clock speed. The register has a known value of 0x00 during reset to ensure that reads from flash will succeed at the fastest process corners, and can be set in one of two ways:

- via Trim Mode, which is necessary before the QA Chip is programmed for the first time; or
- via the CPU, which is necessary every time the QA Chip is powered up before any flash write or erasure accesses can be carried out.

REGISTERS USED FOR TESTING FLASH

There are a number of registers specifically for testing the flash implementation. A single 32-bit write to an appropriate RAM address allows the setting of any combination of these flash test registers.

RAM consists of 96×32 -bit words, and can be pointed to by any of the standard A_n address registers. A write to a RAM address in the range 97-127 does nothing with the RAM (reads return 0), but a write to a RAM address in the range 0x80-0x87 will write to specific groupings of registers according to the low 3 bits of the RAM address. A 1 in the address bit means the appropriate part of the 32-bit Accumulator value will be written to the appropriate flash test registers. A 0 in the address bit means the register bits will be unaffected.

5 The registers and address bit groupings are listed in Table 352:

Table 12. Flash test registers settable from CPU in RAM address range 0x80-0x87⁵

adr bitSuperscript paranumonly	data bits	name	description
0	0	shadowsOff	0 = shadowing applies (nybble based flash access) 1 = shadowing disabled, 8-bit direct accesses to flash.
	1	hiFlashAdr	Only valid when shadowsOff = 1 0 = accesses are to lower 4Kbytes of flash 1 = accesses are to upper 4 Kbytes of flash
	2		
1	3	enableFlashTe st	0 = keep flash test register within the TSMC flash IP in its reset state 1 = enable flash test register to take on non-reset values.
	8-4	flashTest	Internal 5-bit flash test register within the TSMC flash IP (SFC008_08B9_HE). If this is written with 0x1E, then subsequent writes will be according to the TSMC write test mode. You must write a non-0x1E value or reset the register to exit this mode.
2	28-9	flashTime	When timerSel is 1, this value is used for the duration of the program cycle within a standard flash write or erasure. 1 unit = 16 clock cycles (16 \times 100ns typical). Regardless of timerSel, this value is also used for the timeout following power down detection before the QA Chip resets itself. 1 unit = 1 clock cycle (= 100ns typical).

⁵ This is from the programmer's perspective. Addresses sent from the CPU are byte aligned, so the MRU needs to test bit $n+2$. Similarly, checking DRAM address > 128 means testing bit 7 of the address in the CPU, and bit 9 in the MRU.

			<i>Note that this means the programmer should set this to an appropriate value (e.g. 5 µs), just as the localId needs to be set.</i>
	29	timerSel	0 = use internal (default) timings for flash writes & erasures 1 = use flashTime for flash writes and erasures

When none of the address register bits 0-2 are set (e.g. a write to RAM address 0x80), then invalid writes will clear the `illChip` and `retryCount` registers.

For example, set the `A0` register to be 0x80 in RAM. A write to `(A0),0` will write to none of the flash test registers, but will clear the `illChip` and `retryCount` registers. A write to `(A0),7` will write to all of the flash test registers. A write to `(A0),2` will write to the `enableFlashTest` and `flashTest` registers only. A write to `(A0),4` will write to the `flashTime` and `timerSel` registers etc.

Finally, a write to address 0x88 in RAM will cause a device erasure. If `infoBlockSel` is 0, then the device erasure will only be of main memory. If `infoBlockSel` is 1, then the device erasure is of both main memory and the information block (which will also clear the `ChipMask` and the `Fuse`).

Reads of invalid RAM areas will reveal information as follows:

- all invalid addresses in RAM (e.g. 0x80) will return the `illChip` flag in the low bit (`illChip` is set whenever 16 consecutive bad reads occur for a single byte in memory)
- all invalid addresses in RAM with the low address bit set (e.g. 0x81, or `(A0),1` when `A0` holds 0x80), will additionally return the most recent `retryCount` setting (only updated by the chip when a bad read occurs). i.e. bit 0 = `illChip`, bits 4-1 = `retryCount`.

REGISTER SUMMARY

Table 13 provides a summary of the registers used in the CPU.

Table 13. Register summary

Register name	Description	#bits
<code>A[0-3]</code>	address registers	49 =36
<code>Acc</code>	Accumulator	32
<code>C1</code>	general purpose counter and index	3
<code>C2</code>	general purpose counter and index	6
<code>illChip</code>	gets set whenever more than 15 consecutive bad reads from flash occurred (and any program executing has hung)	1
<code>InByte</code>	input byte from outside world	8
<code>Go</code>	determines whether CPU is executing	1
<code>LocalId</code>	determines id for this chip's IO	7
<code>OutByte</code>	output byte to outside world	8
<code>Z</code>	zero flag for last xfer to Acc	1

PCA	program counter array	1212=144
PCRamSel	Program code is executing in flash (0) or ram (1)	1
RetryCount	counts the number of retries for bad reads	4
RTMP	bit used to allow multi-word rotations	1
SP	stack pointer into PCA	4
Trim	trims ring oscillator frequency	8
flash test registers	various registers in the embedded flash and flash access logic specifically for testing the flash memory	30
	TOTAL (bits)	295

STARTUP

Whenever the chip is powered up, or receives a 'write' command over the serial interface, the PC and PCRamSel get set to 0 and execution begins at 0 in Flash memory. The program (starting at 0) needs to

5 determine how the program was started by reading the InByte register.

If the first byte read is 0xFF, the chip is being requested to perform software reset tasks. Execution of software reset can only be interrupted by a power down. The reset tasks include setting up RAM to contain known startup state information, setting up Trim and localID registers etc. The CPU signals that it is now ready to receive commands from an external device by writing to the OutByte register. An external

10 Master is able to read the OutByte (and any further outbytes that the CPU decides to send) if it so wishes by a read using the localId.

Otherwise the first byte read will be of the form where the least significant bit is 0, and bits 7-1 contain the localId of the device as read over the serial interface. This byte is usually discarded since it nominally only has a value of differentiation against a software reset request. The second and subsequent bytes contain

15 the data message of a write using the localId. The CPU can prevent interruption during execution by writing 0 to the localId and then restoring the desired localId at the later stage.

IMPLEMENTATION

The high-level definition of a CPU capable of implementing the functionality required of an QA Chip is as follows.

20 PHYSICAL INTERFACE

Pin connections

The pin connections are described in Table 14.

Table 14. Pin connections to QA Chip

pin	direction	Description
Vdd	In	Nominal voltage. If the voltage deviates from this by more than a fixed amount, the chip will RESET.
GND	In	
SClk	In	Serial clock
SDa	In/Out	Serial data

The system operating clock SysClk is different to SClk. SysClk is derived from an internal ring oscillator based on the process technology. In the FPGA implementation SysClk is obtained via a 5th pin.

Size and cost

The QA Chip uses a 0.25 μm CMOS Flash process for an area of 1mm² yielding a 10 cent manufacturing

5 cost in 2002. A breakdown of area is listed in Table 15.

Tabel 15. Breakdown of Area for QA Chip

approximate area (mm ²)	Description
0.49	8KByte flash memory TSMC: SFC0008_08B9_HE (8K x 8-bits, erase page size = 512 bytes) Area = 724.688 μm x 682.05 μm .
0.08	3072 bits of static RAM
0.38	General logic
0.05	Analog circuitry
1	TOTAL (approximate)

Note that there is no specific test circuitry (scan chains or BIST) within the QA Chip, so the total transistor

10 count is as shown in Table 15.

Reset

The chip performs a RESET upon power-up. In addition, tamper detection and prevention circuitry in the chip will cause the chip to either RESET or erase Flash memory (depending on the attack detected) if an attack is detected.

15 OPERATING SPEED

The base operating system clock SysClk is generated internally from a ring oscillator (process dependant). Since the frequency varies with operating temperature and voltage, the clock is passed through a temperature-based clock filter before use. The frequency is built into the chip during manufacture, and cannot be changed. The frequency is in the range 7-14 MHz.

20 GENERAL MANUFACTURING COMMENTS

Manufacturing comments are not normally made when normally describing the architecture of a chip. However, in the case of the QA Chip, the physical implementation of the chip is very much tied to the security of the key. Consequently a number of specialized circuits and components are necessary for implementation of the QA Chip. They are listed here.

25

- Flash process
- Internal randomized clock
- Temperature based clock filter
- Noise generator
- Tamper Prevention and Detection circuitry

30

- Protected memory with tamper detection
- Boot-strap circuitry for loading program code

- Data connections in polysilicon layers where possible
- OverUnderPower Detection Unit
- No scan-chains or BIST

Flash process

5 The QA Chip is implemented with a standard Flash manufacturing process. It is important that a Flash process be used to ensure that good endurance is achieved (parts of the Flash memory can be erased/written many times).

Internal randomized clock

To prevent clock glitching and external clock-based attacks, the operating clock of the chip should be

10 generated internally. This can be conveniently accomplished by an internal ring oscillator. The length of the ring depends on the process used for manufacturing the chip.

Due to process and temperature variations, the clock needs to be trimmed to bring it into a range usable for timing of Flash memory writes and erases.

The internal clock should also contain a small amount of randomization to prevent attacks where light

15 emissions from switching events are captured, as described below.

Finally, the generated clock must be passed through a temperature-based clock filter before being used by the rest of the chip.

The normal situation for FET implementation for the case of a CMOS inverter (which involves a pMOS transistor combined with an nMOS transistor) as shown in Figure 26.

20 During the transition, there is a small period of time where both the nMOS transistor and the pMOS transistor have an intermediate resistance. The resultant power-ground short circuit causes a temporary increase in the current, and in fact accounts for around 20% of current consumed by a CMOS device. A small amount of infrared light is emitted during the short circuit, and can be viewed through the silicon substrate (silicon is transparent to infrared light). A small amount of light is also emitted during the

25 charging and discharging of the transistor gate capacitance and transmission line capacitance.

For circuitry that manipulates secret key information, such information must be kept hidden.

Fortunately, IBM's PICA system and LVP (laser voltage probe) both have a requirement for repeatability due to the fact that the photo emissions are extremely weak (one photon requires more than 10^5 switching events). PICA requires around 10^9 passes to build a picture of the optical waveform. Similarly the

30 LVP requires multiple passes to ensure an adequate SNR.

Randomizing the clock stops repeatability (from the point of view of collecting information about the same position in time), and therefore reduces the possibility of this attack.

Temperature based clock filter

The QA Chip circuitry is designed to operate within a specific clock speed range. Although the clock is

35 generated by an internal ring oscillator, the speed varies with temperature and power. Since the user supplies the temperature and power, it is possible for an attacker to attempt to introduce race-conditions in the circuitry at specific times during processing. An example of this is where a low temperature causes a clock speed higher than the circuitry is designed for, and this may prevent an XOR from working properly, and of the two inputs, the first may always be returned. These styles of transient fault attacks

40 are documented further in [1]. The lesson to be learned from this is that the input power and operating temperature *cannot be trusted*.

Since the chip contains a specific power filter, we must also filter the clock. This can be achieved with a temperature sensor that allows the clock pulses through only when the temperature range is such that the chip can function correctly.

The filtered clock signal would be further divided internally as required.

5 Noise Generator

Each QA Chip should contain a noise generator that generates continuous circuit noise. The noise will interfere with other electromagnetic emissions from the chip's regular activities and add noise to the I_{dd} signal. Placement of the noise generator is not an issue on an QA Chip due to the length of the emission wavelengths.

10 The noise generator is used to generate electronic noise, multiple state changes each clock cycle, and as a source of pseudo-random bits for the Tamper Prevention and Detection circuitry.

A simple implementation of a noise generator is a 64-bit maximal period LFSR seeded with a non-zero number.

Tamper Prevention and Detection circuitry

15 A set of circuits is required to test for and prevent physical attacks on the QA Chip. However what is actually detected as an attack may not be an intentional physical attack. It is therefore important to distinguish between these two types of attacks in an QA Chip:

- where you *can be certain* that a physical attack has occurred.
- where you *cannot* be certain that a physical attack has occurred.

20 The two types of detection differ in what is performed as a result of the detection. In the first case, where the circuitry can be certain that a true physical attack has occurred, erasure of flash memory key information is a sensible action. In the second case, where the circuitry cannot be sure if an attack has occurred, there is still certainly something wrong. Action must be taken, but the action should not be the erasure of secret key information. A suitable action to take in the second case is a chip RESET. If what

25 was detected was an attack that has permanently damaged the chip, the same conditions will occur next time and the chip will RESET again. If, on the other hand, what was detected was part of the normal operating environment of the chip, a RESET will not harm the key.

A good example of an event that circuitry cannot have knowledge about, is a power glitch. The glitch may be an intentional attack, attempting to reveal information about the key. It may, however, be the result of a

30 faulty connection, or simply the start of a power-down sequence. It is therefore best to only RESET the chip, and not erase the key. If the chip was powering down, nothing is lost. If the System is faulty, repeated RESETs will cause the consumer to get the System repaired. In both cases the consumable is still intact.

A good example of an event that circuitry can have knowledge about, is the cutting of a data line within

35 the chip. If this attack is somehow detected, it could only be a result of a faulty chip (manufacturing defect) or an attack. In either case, the erasure of the secret information is a sensible step to take.

Consequently each QA Chip should have 2 Tamper Detection Lines - one for definite attacks, and one for possible attacks. Connected to these Tamper Detection Lines would be a number of Tamper Detection test units, each testing for different forms of tampering. *In addition, we want to ensure that the Tamper*

40 *Detection Lines and Circuits themselves cannot also be tampered with.*

At one end of the Tamper Detection Line is a source of pseudo-random bits (clocking at high speed compared to the general operating circuitry). The Noise Generator circuit described above is an adequate

source. The generated bits pass through two different paths - one carries the original data, and the other carries the inverse of the data. The wires carrying these bits are in the layer above the general chip circuitry (for example, the memory, the key manipulation circuitry etc.). The wires must also cover the random bit generator. The bits are recombined at a number of places via an XOR gate. If the bits are

5 different (they should be), a 1 is output, and used by the particular unit (for example, each output bit from a memory read should be ANDed with this bit value). The lines finally come together at the Flash memory Erase circuit, where a complete erasure is triggered by a 0 from the XOR. Attached to the line is a number of triggers, each detecting a physical attack on the chip. Each trigger has an oversize nMOS transistor attached to GND. The Tamper Detection Line physically goes through this nMOS transistor. If 10 the test fails, the trigger causes the Tamper Detect Line to become 0. The XOR test will therefore fail on either this clock cycle or the next one (on average), thus RESETing or erasing the chip.

Figure 22 illustrates the basic principle of a Tamper Detection Line in terms of tests and the XOR connected to either the Erase or RESET circuitry.

15 The Tamper Detection Line must go through the drain of an output transistor for each test, as illustrated by Figure 23.

It is not possible to break the Tamper Detect Line since this would stop the flow of 1s and 0s from the random source. The XOR tests would therefore fail. As the Tamper Detect Line physically passes through each test, it is not possible to eliminate any particular test without breaking the Tamper Detect Line.

20 It is important that the XORs take values from a variety of places along the Tamper Detect Lines in order to reduce the chances of an attack. Figure 24 illustrates the taking of multiple XORs from the Tamper Detect Line to be used in the different parts of the chip. Each of these XORs can be considered to be generating a ChipOK bit that can be used within each unit or sub-unit.

A typical usage would be to have an OK bit in each unit that is ANDed with a given ChipOK bit each cycle. The OK bit is loaded with 1 on a RESET. If OK is 0, that unit will fail until the next RESET. If the

25 Tamper Detect Line is functioning correctly, the chip will either RESET or erase all key information. If the RESET or erase circuitry has been destroyed, then this unit will not function, thus thwarting an attacker. The destination of the RESET and Erase line and associated circuitry is very context sensitive. It needs to be protected in much the same way as the individual tamper tests. There is no point generating a RESET pulse if the attacker can simply cut the wire leading to the RESET circuitry. The actual implementation will 30 depend very much on what is to be cleared at RESET, and how those items are cleared.

Finally, Figure 25 shows how the Tamper Lines cover the noise generator circuitry of the chip. The generator and NOT gate are on one level, while the Tamper Detect Lines run on a level above the generator.

Protected memory with tamper detection

35 It is not enough to simply store secret information or program code in flash memory. The Flash memory and RAM must be protected from an attacker who would attempt to modify (or set) a particular bit of program code or key information. The mechanism used must conform to being used in the Tamper Detection Circuitry (described above).

40 The first part of the solution is to ensure that the Tamper Detection Line passes directly above each flash or RAM bit. This ensures that an attacker cannot probe the contents of flash or RAM. A breach of the covering wire is a break in the Tamper Detection Line. The breach causes the Erase signal to be set, thus

deleting any contents of the memory. The high frequency noise on the Tamper Detection Line also obscures passive observation.

The second part of the solution for flash is to always store the data with its inverse. In each byte, 4 bits contains the data, and 4 bits (the shadow) contains the inverse of the data. If both are 0, this is a valid

5 erase state, and the value is 0. Otherwise, the memory is only valid if the 4 bits of shadow are the inverse of the main 4 bits. The reasoning is that it is possible to add electrons to flash via a FIB, but not take electrons away. If it is possible to change a 0 to 1 for example, it is not possible to do the same to its inverse, and therefore regardless of the sense of flash, an attack can be detected.

The second part of the solution for RAM is to use a parity bit. The data part of the register can be checked 10 against the parity bit (which will not match after an attack).

The bits coming from Flash and RAM can therefore be validated by a number of test units (one per bit) connected to the common Tamper Detection Line. The Tamper Detection circuitry would be the first circuitry the data passes through (thus stopping an attacker from cutting the data lines).

In addition, the data and program code should be stored in different locations for each chip, so an 15 attacker does not know where to launch an attack. Finally, XORing the data coming in and going to Flash with a random number that varies for each chip means that the attacker cannot learn anything about the key by setting or clearing an individual bit that has a probability of being the key (the inverse of the key must also be stored somewhere in flash).

Finally, each time the chip is called, every flash location is read before performing any program code.

20 This allows the flash tamper detection to be activated in a common spot instead of when the data is actually used or program code executed. This reduces the ability of an attacker to know exactly what was written to.

Boot-strap circuitry for loading program code

Program code should be kept in protected flash instead of ROM, since ROM is subject to being altered in 25 a non-testable way. A boot-strap mechanism is therefore required to load the program code into flash memory (flash memory is in an indeterminate state after manufacture).

The boot-strap circuitry must not be in a ROM - a small state-machine suffices. Otherwise the boot code could be trivially modified in an undetectable way.

The boot-strap circuitry must erase all flash memory, check to ensure the erasure worked, and then load 30 the program code.

The program code should only be executed once the flash program memory has been validated via Program Mode.

Once the final program has been loaded, a fuse can be blown to prevent further programming of the chip. Connections in polysilicon layers where possible

35 Wherever possible, the connections along which the key or secret data flows, should be made in the polysilicon layers. Where necessary, they can be in metal 1, but must never be in the top metal layer (containing the Tamper Detection Lines).

OverUnder Power Detection Unit

Each QA Chip requires an OverUnder Power Detection Unit (PDU) to prevent Power Supply Attacks. A

40 PDU detects power glitches and tests the power level against a Voltage Reference to ensure it is within a certain tolerance. The Unit contains a single Voltage Reference and two comparators. The PDU would be connected into the RESET Tamper Detection Line, thus causing a RESET when triggered.

A side effect of the PDU is that as the voltage drops during a power-down, a RESET is triggered, thus erasing any work registers.

No scan chains or BIST

Test hardware on an QA Chip could very easily introduce vulnerabilities. In addition, due to the small size

5 of the QA Chip logic, test hardware such as scan paths and BIST units could in fact take a sizeable chunk of the final chip, lowering yield and causing a situation where an error in the test hardware causes the chip to be unusable. As a result, the QA Chip should not contain any BIST or scan paths. Instead, the program memory must first be validated via the Program Mode mechanism, and then a series of program tests run to verify the remaining parts of the chip.

10 Architecture

Figure 28 shows a high level block diagram of the QA Chip. Note that the tamper prevention and detection circuitry is not shown.

ANALOGUE UNIT

Figure 29 shows a block diagram of the Analogue Unit. Blocks shown in yellow provide additional 15 protection against physical and electrical attack and, depending on the level of security required, may optionally be implemented.

Ring oscillator

The operating clock of the chip (SysClk) is generated by an internal ring oscillator whose frequency can be 20 trimmed to reduce the variation from 4:1 (due to process and temperature) down to 2:1 (temperature variations only) in order to satisfy the timing requirements of the Flash memory.

The length of the ring depends on the process used for manufacturing the chip. A nominal operating frequency range of 10 MHz is sufficient. This clock should contain a small amount of randomization to prevent attacks where light emissions from switching events are captured.

Note that this is different to the input scik which is the serial clock for external communication.

25 The ring oscillator is covered by both Tamper Detection and Prevention lines so that if an attacker attempts to tamper with the unit, the chip will either RESET or erase all secret information.

FPGA Note: the FPGA does not have an internal ring oscillator. An additional pin (SysClk) is used instead. This is replaced by an internal ring oscillator in the final ASIC.

Voltage reference

30 The voltage reference block maintains an output which is substantially independant of process, supply voltage and temperature. It provides a reference voltage which is used by the PDU and a reference current to stabilise the ring oscillator. It may also be used as part of the temperature based clock filter.

OverUnder power detection unit (PDU)

35 The Under Voltage Detection Unit provides the signal PwrFailing which, if asserted, indicates that the power supply may be turning off. This signal is used to rapidly terminate any Flash write that may be in progress to avoid accidentally writing to an indeterminate memory location.

Note that the PDU triggers the RESET Tamper Detection Line only. It does not trigger the Erase Tamper Detection Line.

40 The PDU can be implemented with regular CMOS, since the key does not pass through this unit. It does not have to be implemented with non-flashing CMOS.

The PDU is covered by both Tamper Detection and Prevention lines so that if an attacker attempts to tamper with the unit, the chip will either RESET or erase all secret information.

Power-on Reset and Tamper Detect Unit

The Power-on Reset unit (POR) detects a power-on condition and generates the PORstL signal that is fed

5 to all the validation units, including the two inside the Tamper Detect Unit (TDU).

All other logic is connected to RstL, which is the PORstL gated by the VAL unit attached to the Reset tamper detection lines within the TDU. Therefore, if the Reset tamper line is asserted, the validation will drive RstL low, *and can only be cleared by a power-down*. If the tamper line is not asserted, then RstL = PORstL.

10 The TDU contains a second VAL unit attached to the Erase tamper detection lines within the TDU. It produces a TamperEraseOK signal that is output to the MIU (1 = the tamper lines are all OK, 0 = force an erasure of Flash).

Noise generator

15 The Noise Generator (NG) is based on a 64-bit maximal period LFSR loaded with a set non-zero bit pattern on RESET.

The NG must be protected by both Tamper Detection and Prevention lines so that if an attacker attempts to tamper with the unit, the chip will either RESET or erase all secret information.

In addition, the bits in the LFSR must be validated to ensure they have not been tampered with (i.e. a

20 parity check). If the parity check fails, the Erase Tamper Detection Line is triggered.

Finally, all 64 bits of the NG are ORed into a single bit. If this bit is 0, the Erase Tamper Detection Line is triggered. This is because 0 is an invalid state for an LFSR.

TRIM UNIT

The 8-bit Trim register within the Trim Unit has a reset value of 0x00 (to enable the flash reads to succeed

25 even in the fastest process corners), and is written to either by the PMU during Trim Mode or by the CPU in Active Mode. Note that the CPU is only able to write *once* to the Trim register between power-on-reset due to the TrimDone flag which provides overloading of LocalldWE.

The reset value of Trim (0) means that the chip has a nominal frequency of 2.7MHz - 10MHz. The upper of the range is when we cannot trim it lower than this (or we could allow some spread on the acceptable

30 trimmed frequency but this will reduce our tolerance to ageing, voltage and temperature which is the range 7MHz to 14MHz). The 2.7MHz value is determined by a chip whose oscillator runs at 10MHz when the trim register is set to its maximum value, so then it must run at 2.7MHz when trim = 0. This is based on the non-linear frequency-current characteristic of the oscillator. Chips found outside of these limits will be rejected.

35 The frequency of the ring oscillator is measured by counting cycles⁶, in the PMU, over the byte period of the serial interface. The frequency of the serial clock, SClk, and therefore the byte period will be accurately controlled during the measurement. The cycle count (Fmeas) at the end of the period is read over the serial bus and the Trim register updated (Trimval) from its power on default (POD) value. The

⁶Note that the PMU counts using 12-bits, saturates at 0xFFFF, and returns the cycle count divided by 2 as an 8-bit value. This means that multiple measure-read-trim cycles may be necessary to resolve any ambiguity. In any case, multiple cycles are necessary to test the correctness of the trim circuitry during manufacture test.

steps are shown in Figure 30. Multiple measure - read - trim cycles are possible to improve the accuracy of the trim procedure.

A single byte for both Fmeas and Trimval provide sufficient accuracy for measurement and trimming of the frequency. If the bus operates at 400kHz, a byte (8 bits) can be sent in 20 μ s. By dividing the

5 maximum oscillator frequency, expected to be 20MHz, by 2 results in a cycle count of 200 and 50 for the minimum frequency of 5MHz resulting in a worst case accuracy of 2%.

Figure 31 shows a block diagram of the Trim Unit:

The 8-bit Trim value is used in the analog Trim Block to adjust the frequency of the ring oscillator by controlling its bias current. The two lsbs are used as a voltage trim, and the 6 msbs are used as a

10 frequency trim.

The analog Trim Clock circuit also contains a Temperature filter.

IO UNIT

The QA Chip acts as a *slave* device, accepting serial data from an external master via the IO Unit (IOU).

Although the IOU actually transmits data over a 1-bit line, the data is always transmitted and received in

15 1-byte chunks.

The IOU receives commands from the master to place it in a specific operating mode, which is one of:

- **Idle Mode:** is the startup mode for the IOU if the fuse has not yet been blown. *Idle Mode* is the mode where the QA Chip is waiting for the next command from the master. Input signals from the CPU are ignored.
- **Program Mode:** is where the QA Chip erases all currently stored data in the Flash memory (program and secret key information) and then allows new data to be written to the Flash. The IOU stays in *Program Mode* until told to enter another mode.
- **Active Mode:** is the startup mode for the IOU if the fuse has been blown (the program is safe to run). Active Mode is where the QA Chip allows the program code to be executed to process the master's specific command. The IOU returns to *Idle Mode* automatically when the command has been processed, or if the time taken between consuming input bytes (while the master is writing the data) or generating output bytes (while the master is reading the results) is too great.
- **Trim Mode:** is where the QA Chip allows the generation and setting of a trim value to be used on the internal ring oscillator clock value. This must be done for safety reasons before a program can be stored in the Flash memory.

CENTRAL PROCESSING UNIT

The Central Processing Unit (CPU) block provides the majority of the circuitry of the 4-bit microprocessor.

Figure 32 shows a high level view of the block.

MEMORY INTERFACE UNIT

35 The Memory Interface Unit (MIU) provides the interface to flash and RAM. The MIU contains a Program Mode Unit that allows flash memory to be loaded via the IOU, a Memory Request Unit that maps 8-bit and 32-bit requests into multiple byte based requests, and a Memory Access Unit that generates read/write strobes for individual accesses to the memory.

Figure 33 shows a high level view of the MIU block.

MEMORY COMPONENTS

The Memory Components block isolates the memory implementation from the rest of the QA Chip.

The entire contents of the Memory Components block must be protected from tampering. Therefore the logic must be covered by both Tamper Detection Lines. This is to ensure that program code, keys, and

5 intermediate data values cannot be changed by an attacker. The 8-bit wide RAM also needs to be parity checked.

Figure 34 shows a high level view of the Memory Components block. It consists of 8KBytes of flash memory and 3072 bits of parity checked RAM.

RAM

10 The RAM block is shown here as a simple 96×32 -bit RAM (plus parity included for verification). The parity bit is generated during the write.

The RAM is in an unknown state after RESET, so program code cannot rely on RAM being 0 at startup.

The initial version of the ASIC has the RAM implemented by Artisan component RA1SH (96×32 -bit RAM without parity). Note that the RAMOutEn port is active low i.e. when 0, the RAM is enabled, and when 1, the

15 RAM is disabled.

Flash memory

A single Flash memory block is used to hold all non-volatile data. This includes program code and variables. The Flash memory block is implemented by TSMC component SFC0008_08B9_HE [4], which has the following characteristics:

20

- $8K \times 8$ -bit main memory, plus 128×8 -bit information memory
- 512 byte page erase
- Endurance of 20,000 cycles (min)
- Greater than 100 years data retention at room temperature
- Access time: 20 ns (max)

25

- Byte write time: $20\mu s$ (min)
- Page erase time: 20ms (min)
- Device erase time: 200 ms (min)
- Area of 0.494mm^2 ($724.66\mu\text{m} \times 682.05\mu\text{m}$)

The FlashCtrl line are the various inputs on the SFC0008_08B9_HE required to read and write bytes, erase 30 pages and erase the device. A total of 9 bits are required (see [4] for more information).

Flash values are unchanged by a RESET. After manufacture, the Flash contents must be considered to be garbage. After an erasure, the Flash contents in the SFC0008_08B9_HE is all 1s.

VAL blocks

The two VAL units are validation units connected to the Tamper Prevention and Detection circuitry, each 35 with an OK bit. The OK bit is set to 1 on PORstL, and ORed with the ChipOK values from both Tamper Detection Lines each cycle. The OK bit is ANDed with each data bit that passes through the unit.

In the case of VAL₁, the effective byte output from the flash will always be 0 if the chip has been tampered with. This will cause shadow tests to fail, program code will not execute, and the chip will hang.

In the case of VAL₂, the effective byte from RAM will always be 0 if the chip has been tampered with, thus 40 resulting in no temporary storage for use by an attacker.

I/O Unit

The I/O Unit (IOU) is responsible for providing the physical implementation of the logical interface, moving between the various modes (Idle, Program, Trim and Active) according to commands sent by the master.

The IOU therefore contains the circuitry for communicating externally with the external world via the SCiK

5 and SDa pins. The IOU sends and receives data in 8-bit chunks. Data is sent serially, most significant bit (bit 7) first through to least significant bit (bit 0) last. When a master sends a command to an QA Chip, the command commences with a single byte containing an id in bits 7-1, and a read/write sense in bit 0, as shown in Figure 35.

The IOU recognizes a global id of 0x00 and a local id of LocalId (set after the CPU has executed program

10 code at reset or due to a global id / ActiveMode command on the serial bus). Subsequent bytes contain modal information in the case of global id, and command/data bytes in the case of a match with the local id.

If the master sends data too fast, then the IOU will miss data, since the IOU never holds the bus. The meaning of too fast depends on what is running. In Program Mode, the master must send data a little

15 slower than the time it takes to write the byte to flash (actually written as $2 \times$ 8-bit writes, or $40\mu\text{s}$). In ActiveMode, the master is permitted to send and request data at rates up to 500 KHz.

None of the latches in the IOU need to be parity checked since there is no advantage for an attacker to destroy or modify them.

The IOU outputs 0s and inputs 0s if either of the Tamper Detection Lines is broken. This will only come

20 into effect if an attacker has disabled the RESET and/or erase circuitry, since breaking either Tamper Detection Lines should result in a RESET or the erasure of all Flash memory.

The IOU's InByte, InByteValid, OutByte, and OutByteValid registers are used for communication between the master and the QA Chip. InByte and InByteValid provide the means for clients to pass commands and data to the QA Chip. OutByte and OutByteValid provide the means for the master to read data from the QA Chip.

25 • Reads from InByte should wait until InByteValid is set. InByteValid will remain clear until the master has written the next input byte to the QA Chip. When the IOU is told (by the FEU or MU) that InByte has been read, the IOU clears the InByteValid bit to allow the next byte to be read from the client.

• Writes to OutByte should wait until OutByteValid is clear. Writing OutByte sets the OutByteValid bit to signify that data is available to be transmitted to the master. OutByteValid will then remain set until the master has read the data from OutByte. If the master requests a byte but OutByteValid is clear, the IOU sends a NAck to indicate the data is not yet ready.

When the chip is reset via RstL, the IOU enters ActiveMode to allow the PMU to run to load the fuse. Once the fuse has been loaded (when MIUAvail transitions from 0 to 1) the IOU checks to see if the program is

35 known to be safe. If it is not safe, the IOU reverts to IdleMode. If it is safe (FuseBlown = 1), the IOU stays in ActiveMode to allow the program to load up the localId and do any other reset initialization, and will not process any further serial commands until the CPU has written a byte to the OutByte register (which may be read or not at the discretion of the master using a localId read). In both cases the master is then able to send commands to the QA Chip.

40 Figure 36 shows a block diagram of the IOU.

With regards to `InByteValid` inputs, set has priority over reset, although both set and reset in correct operation should never be asserted at the same time. With regards to `IOSetInByte` and `IOLoadInByte`, if `IOSetInByte` is asserted, it will set `InByte` to be `0xFF` regardless of the setting of `IOLoadInByte`.

The two `VAL` units are validation units connected to the Tamper Prevention and Detection circuitry, each with an `OK` bit. The `OK` bit is set to 1 on `PORstL`, and ORed with the `ChipOK` values from both Tamper Detection Lines each cycle. The `OK` bit is ANDed with each data bit that passes through the unit.

In the case of `VAL1`, the effective byte output from the chip will always be 0 if the chip has been tampered with. Thus no useful output can be generated by an attacker. In the case of `VAL2`, the effective byte input to the chip will always be 0 if the chip has been tampered with. Thus no useful input can be chosen by an attacker.

There is no need to verify the registers in the `IOU` since an attacker does not gain anything by destroying or modifying them.

The current mode of the `IOU` is output as a 2-bit `IOMode` to allow the other units within the `QA Chip` to take correct action. `IOMode` is defined as shown in Table 16:

15 Table 16. `IOMode` values

Value	Interpretation
00	Idle Mode
01	Program Mode
10	Active Mode
11	Trim Mode

The Logic blocks generate a 1 if the current `IOMode` is in Program Mode, Active Mode or Trim Mode respectively. The logic blocks are:

20

<code>Logic₁</code>	<code>IOMode</code> = 01 (Program)
<code>Logic₂</code>	<code>IOMode</code> = 10 (Active)
<code>Logic₃</code>	<code>IOMode</code> = 11 (Trim)

STATE MACHINE

There are two state machines in the `IOU` running in parallel. The first is a byte-oriented state machine, the second is a bit-oriented state machine. The byte-oriented state machine keeps track of the operating

25 mode of the `QA Chip` while the bit-oriented state machine keeps track of the low-level bit Rx/Tx protocol. The `SDa` and `SClk` lines are connected to the respective pads on the `QA Chip`. The `IOU` passes each of the signals from the pads through 2 D-types to compensate for metastability on input, and then a further latch and comparitor to ensure that signals are only used if stable for 2 consecutive internal clock cycles.

Start/Stop control signals

30 The `StartDetected` and `StopDetected` control signals are generated based upon monitoring `SDa` synchronized to `SClk`. The `StartDetected` condition is asserted on the falling edge of `SDa` synchronized to `SClk`, and the `StopDetected` condition is asserted on the rising edge of `SDa` synchronized to `SClk`.

In addition we generate `feSClk` which is asserted on the falling edge of `SClk`, and `reSClk` which is asserted on the rising edge of `SClk`. Finally, `feSClkPrev` is the value of `feSClk` delayed by a single cycle.

Figure 37 shows the relationship of inputs and the generation of SDaReg, reSclk, feSclk, feSclkPrev, StartDetected and StopDetected.

The SDaRegSelect logic compensates for the 2:1 variation in clock frequency. It uses the length of the high period of the Sclk (from the saturating counter) to select between sda5, sda6 and sda7 as the valid data from 5 300ns before the falling edge of Sclk as follows.

The minimum time for the high period of Sclk is 600ns. If the counter ≤ 4 (i.e. 5 or fewer cycles with Sclk = 1) then SDaReg output = sda5 (sample point is equidistant from rising and falling edges). If the counter = 5 or 6 (i.e. 6 or 7 samples where Sclk = 1), then SDaReg output = sda6. If the counter = 7 (the counter saturates when there are 8 samples of Sclk = 1), then SDaReg output = sda7. This is shown in pseudocode 10 below:

```
If ((counter2 = 0)  $\vee$  (counter = 4))  
    SDaReg = sda5  
ElseIf (counter = 7)  
    SDaReg = sda7  
Else  
    SDaReg = sda6  
EndIf
```

The counter also provides a means of enabling start and stop detection. There is a minimum of a 600ns setup and 600ns hold time for start and stop conditions. At 14MHz this means samples 4 and 5 after the 20 rising edge (sample 1 is considered to be the first sample where Sclk = 1) could potentially include a valid start or stop condition. At 7 MHz samples 4 and 5 represent 284 and 355ns respectively, although this is after the rising edge of Sclk, which itself is 100ns after the setup of data (i.e. 384 and 455ns respectively and therefore safe for sampling). Thus the data will be stable (although not a start or stop). Since we 25 detect stops and starts using sda5 and sda6, we can only validly detect starts and stops 6 cycles after a rising edge, and we need to not-detect starts and stops 4 cycles before the falling edge. We therefore only detect starts and stops when the counter is ≥ 6 (i.e. when sclk3 and sclk2 are 0 and 1 respectively, sda2 holds sample 1 coincident with the rising edge, sda1 holds sample 2, sda0 holds sample 3, we load the counter with 0 and sample SDa to obtain the new sda0 which will hold sample 4 at the end of the cycle. Thus while the counter is incrementing from 0 to 1, sda0 will hold sample 4. Therefore sample 4 will be in 30 sda6 when the counter is 6.

Control of SDa and Sclk pins

The Sclk line is always driven by the master. The SDa line is driven low whenever we want to transmit an ACK (SDa is active low) or a 0-bit from OutByte. The generation of the SDa pin is shown in the following pseudocode:

```
35      TxAck = (bitSM_state = ack)  $\wedge$  ((byteSM_state = doWrite)  $\vee$   
           (((byteSM_state = getGlobalCmd)  $\vee$  (byteSM_state = checkId))  $\wedge$   
           AckCmd))  
      TxBit  $\leftarrow$  (byteSM_state = doRead)  $\wedge$  (bitSM_state = xferBit)  $\wedge$   
            $\neg$ OutByte_bitCount  
40      SDa =  $\neg$ (TxAck  $\vee$  TxBit) # only drive the line when we are xmitting a 0
```

The slew rate of the SDa line should be restricted to minimise ground bounce. The pad must guarantee a fall time > 20 ns. The rise time will be controlled by the external pull up resistor and bus capacitance.

RING OSCILLATOR

SysClk is required to be in the range 7 - 14 MHz throughout the lifetime of the circuit provided V_{DD} is maintained within the range $V_{DDMIN} < V_{DD} < V_{DDMAX}$. The 2:1 range is derived from the programming time requirements of the TSMC Flash memory. If this range is exceeded, the useful lifetime of the Flash may

5 be reduced.

The first version of the QA chip, without physical protection, does not require the addition of random jitter to the clock. However, it is recommended that the ring oscillator be designed in such a way as to allow for the addition of jitter later on with minimal modification. In this way, the un-trimmed centre frequency would not be expected to change.

10 The initial frequency error must be reduced to remain within the range 10MHz / 1.41 to 10MHz \times 1.41 allowing for variation in:

- voltage
- temperature
- ageing

15 • added jitter

- errors in frequency measurement and setting accuracy

The range budget must be partitioned between these variables.

Figure 38 is a Ring oscillator block diagram

The above arrangement allows the oscillator centre frequency to be trimmed since the bias current of the

20 ring oscillator is controlled by the DAC. SysClk is derived by dividing the oscillator frequency by 5 which makes the oscillator smaller and allows the duty cycle of the clock to be better controlled.

DAC (programmable current source)

Using V_{bg} , this block sources a current that can be programmed by the Trim signal. 6 of the available 8 trim bits will be used (trim₇₋₂) giving a clock adjustment resolution of about 250kHz. The range of current

25 should be such that the ring oscillator frequency can be adjusted over a 4 to 1 range.

Table 17. Programmable current source target performance

Parameter	Conditions	Min	Typ	Max	Units
Iout	Trim7-2 = 0		5		μ A
	Trim7-2 = 32		12.5		
	Trim7-2 = 63		20		
Vrefin			1.23		V
Rout	Trim7-2 = 63	2.5			$M\Omega$

Ring oscillator circuit

30 Table 18. Ring oscillator target performance

Parameter	Conditions	Min	Typ	Max	Units
Fosc ⁷		7	10	14	MHz

⁷ Accounting for division by 5

IDD			10		µA
K _I			1		MHz/µA
K _{VDD}			+200		KHz/V
K _T			+30		KHz/oC
V _{start}		1.5			V

K_I = control sensitivity, $K_{VDD} = V_{DD}$ sensitivity, K_T = temperature sensitivity

With the figures above, K_{VDD} will give rise to a maximum variation of $\pm 50\text{kHz}$ and K_T to $\pm 1.8\text{MHz}$ over the specified range of V_{DD} and temperature.

Div5

5 The ring oscillator will be prescaled by 5 to obtain the nominal 10MHz clock. An asynchronous design may be used to save power. Several divided clock duty cycles are obtainable, eg 4:1, 3:2 etc. To ease timing requirements for the standard cell logic block, the following clock will be generated; most flip-flops will operate on the rising edge of the clock allowing negative edge clocking to meet memory timing.

Table 19. Div5 target performance

10

Parameter	Conditions	Min	Typ	Max	Units
F _{max}	V _{dd} = 1.5V	100			MHz
IDD			10		µA

POWER ON RESET

This block combines the overL (omitted from the current version), underL and MAURstOutL signals to provide the global reset. MAURstOutL is delayed by one clock cycle to ensure a reset generated when this signal is asserted has at least this duration since the reset deasserts the signal itself. It should be noted that the

15 register, with active low reset RN, is the only one in the QA chip not connected to RstL.

[4] TSMC, Oct 1, 2000, *SFC0008_08B9_HE*, 8K × 8 Embedded Flash Memory Specification, Rev 0.1.
 [5] TSMC (design service division), Sep 10, 2001, *0.25um Embedded Flash Test Mode User Guide*, V0.3.
 [6] TSMC (EmbFlash product marketing), Oct 19, 2001, *0.25um Application Note*, V2.2.
 20 [7] Artisan Components, Jan 99, Process Perfect Library Databook 2.5-Volt Standard Cells, Rev1.0.